

Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft

Band 51

Semantische Netze in Informationssystemen
Verbesserung der Suche
durch
Interaktion und Visualisierung

Daniela Trunk

Dezember 2005

Fachhochschule Köln
Fakultät für Informations- und Kommunikationswissenschaften
Institut für Informationswissenschaft

Trunk, Daniela
Semantische Netze in Informationssystemen
Verbesserung der Suche durch Interaktion und Visualisierung
Köln : Fachhochschule Köln,
Fakultät für Informations- und Kommunikationswissenschaften,
Institut für Informationswissenschaft, 2005
(Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; Nr. 51)

ISSN (elektronische Version) 1434-1115

Die Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft berichten über aktuelle Forschungsergebnisse des Instituts Informationswissenschaft der Fachhochschule Köln. Veröffentlicht werden sowohl Arbeiten der Dozentinnen und Dozenten als auch herausragende Arbeiten der Studierenden. Die Kontrolle der wissenschaftlichen Qualität der Veröffentlichungen liegt bei der Schriftleitung. Jeder Band erscheint in elektronischer Version (im WWW: <http://www.fbi.fh-koeln.de/institut/papers/arbeitspapiere.php>).

Fachhochschule Köln
Fakultät für Informations- und Kommunikationswissenschaften
Institut für Informationswissenschaft
Claudiusstr.1 D-50678 Köln
Tel.: 0221/8275-3376, Fax: 0221/3318583
E-Mail: schriftenreihe@fbi.fh-koeln.de

Schriftleitung: Prof. Dr. Konrad Scherfer, Saskia Brintrup
© FH Köln 2005

ABSTRACT

Semantische Netze unterstützen den Suchvorgang im Information Retrieval. Sie bestehen aus relationierten Begriffen und helfen dem Nutzer das richtige Vokabular zur Fragebildung zu finden. Eine leicht und intuitiv erfassbare Darstellung und eine interaktive Bedienungsmöglichkeit optimieren den Suchprozess mit der Begriffsstruktur. Als Interaktionsform bietet sich Hypertext mit dem etablierte Point- und Klickverfahren an. Eine Visualisierung zur Unterstützung kognitiver Fähigkeiten kann durch eine Darstellung der Informationen mit Hilfe von Punkten und Linien erfolgen. Vorgestellt werden die Anwendungsbeispiele Wissensnetz im Brockhaus multimedial, WordSurfer der Firma BiblioMondo, SpiderSearch der Firma BOND und Topic Maps Visualization in dandelon.com und im Portal Informationswissenschaft der Firma AGI – Information Management Consultants.

INHALTSVERZEICHNIS

1	EINLEITUNG	8
2	SEMANTISCHE NETZE IN INFORMATIONSSYSTEMEN	11
2.1	Informationssysteme und Information Retrieval	11
2.2	Repräsentation von Dokumenteninhalten mit Sprache und Suchprozess im Information Retrieval	12
2.2.1	Probleme mit sprachlicher Vielfalt und fehlendem Kontext im Information Retrieval	12
2.2.2	Suchprozess im Information Retrieval	15
2.3	Optimierungsmöglichkeiten	16
2.3.1	Suchfunktionen	16
2.3.2	Linguistische automatische Indexierung	17
2.3.3	Statistische Automatische Indexierung	18
2.4	Semantische Netze	19
2.4.1	Semantische Netze im Information Retrieval	19
2.4.2	Bereitstellung semantischer Netze	21
3	INTERAKTION	27
3.1	Mensch-Computer-Interaktion	27
3.2	Interaktion in semantischen Netzen	30
4	VISUALISIERUNG	33
4.1	Informationsvisualisierung	33
4.2	Visualisierung semantischer Netze	36
5	ANWENDUNGSBEISPIELE	41
5.1	Wissensnetz im Brockhaus multimedial	41
5.2	WordSurfer	52
5.3	SpiderSearch	65
5.4	Topic Maps Visualization in dandelon.com und Portal Informationswissenschaft	79
6	RESÜMEE UND AUSBLICK	102
7	LITERATURVERZEICHNIS	103
7.1	Literatur	103
7.2	Weitere Quellen	110

ABBILDUNGSVERZEICHNIS

Abb. 1	Term-Dokument-Matrix. (Quelle: Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 146)	24
Abb. 2	Term-Term-Matrix (Quelle: Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 147)	24
Abb. 3	Term-Beziehungsmatrix (Quelle: Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 147)	25
Abb. 4	Baumdarstellung in MIDOSThesaurus	37
Abb. 5	Baumdarstellung in H3	38
Abb. 6	Unübersichtliche Netzdarstellung mit vielen Knoten und Linien (Quelle: Däßler: Informationsvisualisierung, 1999, S. 16)	39
Abb. 7	Brockhaus – Anzeige des Artikels „Buch“	42
Abb. 8	Brockhaus – Pop-up-Fenster zum Eintrag „Buch“	43
Abb. 9	Brockhaus – Wissensnetz (Standardeinstellung) zum Eintrag „Buch“	43
Abb. 10	Brockhaus – Wissensnetz (erweiterte Einstellung) zum Eintrag „Buch“	44
Abb. 11	Brockhaus – Vorschau auf ein neues Begriffsnetz	45
Abb. 12	Brockhaus – Wissensnetz mit nicht vollständig angezeigten Einträgen	46
Abb. 13	Brockhaus – Wissensnetz mit vollständig angezeigtem Eintrag	46
Abb. 14	Brockhaus – Anzeige des Artikels zum Synonym „Inkunabeln“	47
Abb. 15	Brockhaus – Anzeige des Artikels „Wiegendrucke“	48
Abb. 16	Brockhaus – Auswahl des Eintrags „Gutenberg: Der Buchdruck und die Folgen“	49
Abb. 17	Brockhaus – Auswahl des Eintrags „Schrift“	49
Abb. 18	Brockhaus – Auswahl des Eintrags „Runen: Geheimnis und Geraune“	50
Abb. 19	Brockhaus – Auswahl des Eintrags „Wikinger“	51
Abb. 20	Brockhaus – Anzeige des Artikels „Wikinger“	52

Abb. 21	N.N. 2005a – Suche nach „Buch“	54
Abb. 22	N.N. 2005b – Auswahl des Eintrags „Veröffentlichung“ im Netz zu „Buch“	55
Abb. 23	N.N. 2005c – Suche nach „Buchdruck“, Auswahl des Eintrags „Druck“	56
Abb. 24	N.N. 2005d – Auswahl des Eintrags „Papier“ im Netz zu „Druck“	57
Abb. 25	N.N. 2005e – Netz zu „Papier“	58
Abb. 26	N.N. 2005f – Navigationsleiste	59
Abb. 27	N.N. 2005g – Suche nach „Bibliothekswesen“, Auswahl des Eintrags „Bibliothekswelt“	59
Abb. 28	N.N. 2005h – Anzeige des Terms „Bibliothekswelt“ ausgewählt aus dem Netz zu „Bibliothekswesen“ (Überlappung)	60
Abb. 29	N.N. 2005i – Suche nach dem Homonym „Krebs“	61
Abb. 30	N.N. 2005j – Suche nach „Öffentliche Bibliothek“	62
Abb. 31	N.N. 2005k – Suche nach „Tanne“	63
Abb. 32	N.N. 2005l – Anzeige des Terms „Buch“ ausgewählt aus dem Netz zu „Bach“	65
Abb. 33	N.N. 2005m – Suche nach „Buch“	67
Abb. 34	N.N. 2005n – Auswahl des Eintrags „Abenteuer“ im Netz zu „Buch“	68
Abb. 35	N.N. 2005o – Animation des Netzübergangs von „Buch“ zu „Abenteuer“	69
Abb. 36	N.N. 2005p – Suche nach „Buchdruck“, Auswahl des Terms „Wissen“	70
Abb. 37	N.N. 2005q – Auswahl des Eintrags „Geschichte“ im Netz zu „Wissen“	71
Abb. 38	N.N. 2005r – Netz zu „Geschichte“	72
Abb. 39	N.N. 2005s – Begriffssammelleiste	73
Abb. 40	N.N. 2005t – Suche nach „Öffentliche Bibliothek“	74
Abb. 41	N.N. 2005u – Auswahl des Eintrags „Instandsetzung“ im Netz zu „Käfer“	75
Abb. 42	N.N. 2005v – Netz zu „Instandsetzung“	76

Abb. 43	N.N. 2005w – Suche nach „VW-Käfer“	77
Abb. 44	N.N. 2005x – Suche nach „Rose“	78
Abb. 45	Portal Informationswissenschaft – Suchoberfläche	81
Abb. 46	Portal Informationswissenschaft – Thesaurusauswahl.....	82
Abb. 47	Portal Informationswissenschaft – Topic Maps Visualization „Bibliothek“	83
Abb. 48	Portal Informationswissenschaft – Navigationsleiste	84
Abb. 49	Portal Informationswissenschaft – Pop-up-Fenster mit Schaltfläche „Zur Suche hinzufügen“	85
Abb. 50	Portal Informationswissenschaft – Pop-up-Fenster zur Hierarchierelation.....	86
Abb. 51	Portal Informationswissenschaft – Auswahl des Eintrags „Allgemeinbibliothek“ im Netz zu „Bibliothek“	88
Abb. 52	Portal Informationswissenschaft – Topic Maps Visualization „Allgemeinbibliothek“	89
Abb. 53	Portal Informationswissenschaft – Topic Maps Visualization „Magazin (Archiv)“	90
Abb. 54	Portal Informationswissenschaft – Thesaurusauswahl zu „Information Retrieval“	91
Abb. 55	Portal Informationswissenschaft – Topic Maps Visualization „Information Retrieval“	92
Abb. 56	Portal Informationswissenschaft – Thesaurusauswahl zu „Blindenbibliothek Hörbibliothek“	93
Abb. 57	Portal Informationswissenschaft – Thesaurusauswahl zu „Blindenbibliothek AND Hörbibliothek“	94
Abb. 58	Portal Informationswissenschaft – Topic Maps Visualization „Dezimalklassifikation“ aus „Experimente mit SWD“	95
Abb. 59	Portal Informationswissenschaft – Topic Maps Visualization „Dezimalklassifikation“ aus „Experimente 2 mit SWD“	96
Abb. 60	Portal Informationswissenschaft – Topic Maps Visualization „Klassifikation / Bibliothek“ aus „Experimente 2 mit SWD“ ..	97
Abb. 61	Portal Informationswissenschaft – Topic Maps Visualization „Buchaufstellung“ aus „Experimente mit SWD“	98

Abb. 62	Portal Informationswissenschaft – Topic Maps Visualization „Buchaufstellung“ aus „Experimente 2 mit SWD“	99
Abb. 63	Portal Informationswissenschaft – Topic Maps Visualization „Thesaurus“ aus „Experimente mit SWD“	100
Abb. 64	Portal Informationswissenschaft – Topic Maps Visualization „Thesaurus“ aus „Experimente 2 mit SWD“	101

ABKÜRZUNGSVERZEICHNIS

IDF	inverse Dokumenthäufigkeit
SWD	Schlagwortnormdatei
RSWK	Regeln für den Schlagwortkatalog
OCR	Optical Character Recognition
OPAC	Online Public Access Catalogue
DGI	Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis
nfd	Information - Wissenschaft und Praxis (Zeitschrift)
WYSIWYG	What You See Is What You Get

1 EINLEITUNG

Seit ein paar Jahren gibt es Informationssysteme auf dem Markt, die den Anspruch haben, eine Suche durch visuelle Darstellungen und interaktive Nutzungsmöglichkeiten intuitiv und leicht handhabbar zu gestalten. Dies wird in der vorliegenden Arbeit aufgegriffen, indem Möglichkeiten zur Optimierung der Suche in Informationssystemen dargestellt werden. Interaktion und Visualisierung werden dabei im Zusammenhang mit semantischen Netzen als Mittel zur Verbesserung des Interfaces betrachtet.

Ausgangspunkt für eine Optimierung der Suche ist zunächst die Beschreibung der Probleme, die sich bei der Repräsentation von Dokumentinhalten mit Sprache ergeben. Daran anschließend werden Verbesserungsmöglichkeiten aufgezeigt, die helfen, diese Schwierigkeiten zu lösen. Neben den standardmäßig eingesetzten Retrievalfunktionen sind dies automatische Indexierungsverfahren und das Einbinden von semantischen Netzen für die Suche. Für den Einsatz von semantischen Netzen in Informationssystemen werden intellektuell erstellte Thesauri und statistisch generierte Begriffsbeziehungen angeführt. Als das umfangreichste terminologisch kontrollierte deutsche Vokabular wird im Rahmen dieser Arbeit die *Schlagwortnormdatei (SWD)* genannt. Die Nutzung der *SWD* bietet sich für Bibliothekssoftware aufgrund ihrer thematischen Universalität und aufgrund der Verfügbarkeit in bibliothekarischen Einrichtungen an. Dabei liegen jedoch Mängel in der Begriffsstruktur vor, die in dieser Arbeit kurz genannt werden. Es folgt dann die Beschreibung von Interaktion und Visualisierung als Mittel zur Verbesserung der Suche. Zunächst erfolgt die Beschreibung verschiedener Interaktionsformen. Für die Interaktion mit semantischen Netzen wird Hypertext und das Prinzip der Navigation und des Browsers vorgestellt. Bei der Betrachtung von Visualisierung wird der Bereich der Informationsvisualisierung erläutert und es wird die Visualisierung semantischer Netze beschrieben. Dabei wird auf unzureichend gelöste Probleme bei der Darstellung vieler Begriffe und Relationen auf einer kleinen Bildschirmfläche hingewiesen. Im letzten Teil der Arbeit werden konkrete Anwendungen vorgestellt, die die Idee von einer möglichen Verbesserung der Suche durch visualisierte semantische Netze anbieten und interaktiv zu bedienen sind. Dies dient einer Bestandsaufnahme und führt nicht zu einer Beurteilung der Qualität der Umsetzung und den tatsächlichen Nutzen für das Retrieval. Kriterien zur Evaluation solcher Anwendungen befinden sich noch im Entste-

hungsprozess und die hier zusammengestellten Beispiele können als Grundlage dienen, solche zu ermitteln.¹

¹ Einige Kriterien zur Evaluierung von Thesaurusstrukturen werden von Sieber vorgestellt. Siehe dazu Sieber, Wolfram: Visualisierung von Thesaurusstrukturen unter besonderer Berücksichtigung eines Hyperbolic Tree Views. Diplomarbeit im Studiengang Informationswirtschaft, 2004

2 SEMANTISCHE NETZE IN INFORMATIONSSYSTEMEN

In diesem Kapitel wird die Suche in elektronischen Informationssystemen und den damit verbundenen Problemen dargestellt. Daran anschließend werden Möglichkeiten zur Optimierung mit besonderem Blickpunkt auf die Bereitstellung von semantischen Netzen vorgestellt. Für die Einbindung von semantischen Netzen in das Retrieval werden die Möglichkeiten von Thesauri und Methoden der statistischen Generierung von Begriffsbeziehungen aufgezeigt.

2.1 Informationssysteme und Information Retrieval

Informationssysteme haben die Funktion eine Suche nach Informationen zu erleichtern.² Ein solches System kann zum Beispiel eine Bibliothek oder ein elektronisches Hilfsmittel, wie eine Literaturdatenbank mit bibliographischen Angaben oder eine Volltextdatenbank sein. Die in dieser Arbeit betrachteten elektronischen Informationssysteme sind Information Retrievalsysteme. Information Retrieval dient dem Auffinden von Dokumenten, wobei ein Information Retrievalsystem das dafür verwendete Werkzeug ist.³ Information Retrieval kann als „inhaltliche Suche in Texten“⁴ beschrieben werden, wenn die Kollektion aus Textdokumenten besteht. Neben weiteren Anwendungen, wie beispielsweise Bildretrieval, ist Textretrieval der wichtigste Bereich des Information Retrievals.⁵ Der Zugriff auf die Dokumente oder auf deren Stellvertreter, wie beispielsweise bibliographische Angaben, erfolgt über eine Suchanfrage.⁶ Dazu muss der Informationsbedarf in Form von einzelnen Wörtern übermittelt werden, die den gesuchten Sachverhalt beschreiben. Relevante Dokumente werden ermittelt, indem die Anfrage vom System mit den gespeicherten Dokumenten verglichen wird.⁷ Die Retrievalfunktion besteht also darin, dass Dokumente bestimmt werden, die zu einer Suchanfrage ähnlich und somit relevant für das Suchinteresse sind.⁸ Zum Abgleich wird nicht auf den Text der Dokumente selbst zugegriffen, sondern auf Indexterme. Dies sind Wörter beziehungsweise Zeichenketten, die dem Dokument entnommen oder aus einer Dokumentationssprache zu-

² Vgl. Salton, Gerard; McGill, Michael: Information Retrieval. Grundlegendes für Informationswissenschaftler, 1987, S. 2

³ Vgl. Ferber, Reginald: Information Retrieval. Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web, 2003, S. 21

⁴ Fuhr, Norbert: Information Retrieval. Skriptum zur Vorlesung im SS 04, S. 6

⁵ Vgl. ebd.

⁶ Vgl. Salton; McGill: Information Retrieval, 1987, S. 8

⁷ Vgl. ebd., S. 11

⁸ Vgl. ebd., S. 12

geteilt wurden und in einer separaten Liste, der invertierten Liste, abgelegt sind.⁹ Standardmäßig erfolgt die Ähnlichkeitsbestimmung zwischen Suchanfrage und Indextermen durch Exact-Match. Das bedeutet, dass nur bei einer genauen Übereinstimmung der Wörter der Suchanfrage mit den Indextermen der invertierten Liste die zugehörigen Dokumente auch als relevant eingestuft und als Treffer ausgegeben werden.

Nach Fuhr sind vage Anfragen und unsicheres Wissen kennzeichnend für Information Retrieval.¹⁰ Die Unsicherheit im Textretrieval ergäbe sich „im wesentlichen durch die beschränkten Möglichkeiten zur Repräsentation der Semantik der natürlichen Sprache“¹¹. Vage Anfragen seien „dadurch gekennzeichnet, daß ihre Antwort nicht a priori definiert ist“¹² und insbesondere dazu Fragen zählen, „die nur iterativ im Dialog durch Reformulierung (in Abhängigkeit von den bisherigen Systemantworten) beantwortet werden können.“¹³

2.2 Repräsentation von Dokumenteninhalten mit Sprache und Suchprozess im Information Retrieval

In den folgenden beiden Kapiteln wird betrachtet, welche Schwierigkeiten es mit der Repräsentation von Dokumenteninhalten mit Sprache und dem Suchprozess im Information Retrieval gibt.

2.2.1 Probleme mit sprachlicher Vielfalt und fehlendem Kontext im Information Retrieval

Die Unsicherheit des gespeicherten Wissens im Zusammenhang mit Retrieval ergibt sich aus den Schwierigkeiten, die bei der Wiedergabe von Inhalten mit Sprache entstehen. Neben einer vielfältigen Ausdrucksmöglichkeit ist es im Retrieval ein Problem, dass Sprache kontextverhaftet ist. Isolierte Wörter, wie sie im Information Retrieval in Suchanfragen und Indexarten vorliegen, können zu Missverständnissen führen, da sie nicht alle Informationen eindeutig wiedergeben. Hinzu kommt, dass Begriffe individuell unterschiedlich gedacht werden und in einem subjektiven semantischen Umfeld eingebettet sind, das ebenfalls für die Suche relevant sein kann.

Sprache hat eine Symbolfunktion und dient der Darstellung von Begriffen.¹⁴ Die Begriffe selbst sind der „Ebene der geistigen Verarbeitung, der

⁹ Vgl. Lepsky, Klaus: Methoden und Verfahren des Information Retrieval. In: Lehrmaterialien zur Veranstaltung. Vorlesungsskript, S. 42

¹⁰ Vgl. Fuhr: Information Retrieval. Skriptum zur Vorlesung im SS 04, S. 7

¹¹ Fuhr, Norbert: Konzepte zur Gestaltung zukünftiger Information-Retrieval-Systeme. In: Experimentelles und praktisches Information Retrieval. Festschrift für Gerhard Lustig, 1992, S. 59-75, hier: S. 68

¹² Ebd., S. 60

¹³ Ebd.

¹⁴ Vgl. Lepsky, Klaus: Inhaltserschließung II. In: Lehrmaterialien zur Veranstaltung. Vorlesungsskript, S. 2

Wahrnehmung und Auseinandersetzung¹⁵ des Menschen mit den Dingen und Objekten der Umwelt zuzuordnen. Um solche Konzepte kommunizieren zu können, werden beispielsweise Wörter einer Sprache verwendet. Da die Begriffsbildung und die Benennung einem individuellen Prozess unterliegen¹⁶, ist es nicht so, „daß in Wörtern der Sprache die Bedeutung von Begriffen kodifiziert ist und diese Bedeutung mit der Sprache abgerufen und rekonstruiert werden kann“¹⁷. Durch die Entwicklung im Laufe eines Lebens lernt der Mensch zwar im Austausch mit anderen, sich über bestimmte Begriffe zu verständigen und Sprache fällt dabei eine normierende Funktion zu, trotzdem bleibt die Bedeutungsgebung und Ausdrucksweise individuell.¹⁸ Gleiche Wörter repräsentieren individuell unterschiedliche Begriffe und gleiche Sachverhalte können je nach Wahrnehmung zu unterschiedlichen Benennungen führen. Für das Information Retrieval bedeutet dies, dass Darstellungen bestimmter Sachverhalte nicht eindeutig und einheitlich beschrieben und Fragen nach gleichen Inhalten unterschiedlich formuliert werden können. Außerdem ist Sprache zum Austausch von Konzepten historisch gewachsen¹⁹ und nicht immer haben sich dabei auch für alle Begriffe Benennungsmöglichkeiten herausgebildet²⁰. Entsprechend vielfältig können Umschreibungen für solche „namenlose“ Sachverhalte sein oder zu neuen Wortkreationen führen. Uneindeutigkeiten ergeben sich auch mit Synonymen, Begriffen, zu denen mehrere Benennungen zur Verfügung stehen, und Quasi-Synonymen, Wörtern, die in einem bestimmten Kontext synonym verwendet werden. Darüber hinaus ist beim Retrieval mit Exact-Match nicht nur die Vokabularwahl der Suchanfrage selbst entscheidend für die Trefferbildung, sondern auch die grammatische Form der Benennungen. Beispielsweise können Substantive im Singular, im Plural oder in flektierten Varianten im Dokument vorliegen und so zu einem negativen Vergleich von Suchanfrage und Text führen. Zusätzlich können mehrere Substantive häufig gleichbedeutend als Adjektiv-Substantiv-Verbindung oder als Kompositum ausgedrückt werden. Umgekehrt kann jedoch nicht jedes Kompositum unter Beibehaltung der gleichen Bedeutung in Substantive zerlegt werden. So sind manche Komposita Individualnamen, wie beispielsweise „Klammeraffe“, und würden bei einer Dekomposition ihre Bedeutung ändern. Zu einer Abweichung zwischen Such- und Indexterm kann es auch kommen, wenn nach Verben und Adjektiven aus dem Dokumententext in substantivierter Form gesucht wird. Des Weiteren drückt sich die Uneindeutigkeit von Sprache auch in Homonymen aus, also in Benennungen, die für verschiedene Begriffe verwendet werden und nur aus dem Kontext heraus eindeutig zu verstehen sind. Werden diese Wörter

¹⁵ Wersig, Gernot: Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis, 1985, S. 13

¹⁶ Vgl. Gödert, Winfried: Sprache und Bedeutung. In: Materialien zur Inhaltserschließung, WS 2003/04, S. 1

¹⁷ Gödert, Winfried: Der konstruktivistische Ansatz für Kommunikation und Informationsverarbeitung. In: Wolfenbütteler Notizen zur Buchgeschichte 27 (2002) 2, S. 199-218, hier: S. 206

¹⁸ Vgl. Gödert: Sprache und Bedeutung, WS 2003/04, S. 1

¹⁹ Vgl. Wersig: Thesaurus-Leitfaden, 1985, S. 13

²⁰ Vgl. Gödert: Sprache und Bedeutung, WS 2003/04, S. 6

deutig zu verstehen sind. Werden diese Wörter isoliert betrachtet, sind sie vieldeutig und führen im Retrieval zu unpräzisen Ergebnissen.

In der Regel kommen Begriffe zum Beschreiben von Inhalten nicht einzeln vor, sondern treten zusammengesetzt zu Themen auf. Um komplexere Sachverhalte wiederzugeben, werden mehrere Begriffe in Beziehung zueinander gesetzt, so dass sich die Bedeutung einzelner Benennungen zur Darstellung von Ereignissen ergänzen.²¹ Die Bezüge zwischen den Begriffen werden dabei durch Syntax hergestellt, so dass Syntax und Semantik im Zusammenspiel Bedeutung schaffen.²² Auch die Begriffsbildung selbst erfolgt beim Menschen nicht isoliert, sondern die Wahrnehmung der Umwelt regt wohl kognitive Verarbeitungsprozesse an, „die für das Erkennen immer auch Vergleich mit bereits Vorhandenem und Einpassen in die vorhandenen Muster beinhalten.“²³ Soll etwas verstanden werden, „muss in der kognitiven Struktur bereits ein Schema vorhanden sein, das den Vergleich mit den eingehenden Sinneseindrücken erlaubt.“²⁴ Neue Begriffe werden durch Anpassen an vorhandene Muster erzeugt.²⁵ Die ganzen Vorgänge des Erkennens beruhen auf einer angeborenen Veranlagung und entwickeln sich im Laufe des Lebens eines Menschen. Dadurch kommt es durch persönliche Erfahrungen und Eindrücke auch zu unterschiedlichen Schemata und zu unterschiedlichen Verarbeitungsprozessen von Sinneseindrücken.²⁶ „Menschen denken nicht in hierarchischen Strukturen, sondern assoziativ“²⁷ und „verknüpfen Informationen in immer wieder neuen Zusammenhängen – abhängig vom Kontext, sprich ihrem individuellen Erfahrungshintergrund sowie der aktuellen Situation oder dem Ziel ihres Handelns.“²⁸ Durch den Austausch mit Mitmenschen werden die individuellen Wahrnehmungen in Teilen einander angeglichen.²⁹ Je mehr dabei auf gleiche Erfahrungen zurückgegriffen wird und eine Beschäftigung mit gleichen Dingen zugrunde liegt, desto mehr verfügt diese Gruppe über gleiche Wissensstrukturen. Und umgekehrt bedeutet dies, je individueller die Erfahrungen sind, beziehungsweise je weniger ein Austausch mit anderen darüber stattfindet, desto individueller sind auch die entwickelten Strukturen, die Begriffsbildung und der wahrgenommene Kontext.³⁰ Durch die kognitiven Vorgänge der Wahrnehmung und Interpretation nehmen die Beziehungen zwischen den Begriffen für die Bedeutungsgebung eine wichtige Funktion ein. Zu den Schwierigkeiten im Retrieval aufgrund unterschiedlicher Benennungsmöglichkeiten kommt hinzu, dass Begriffe nicht nur individuell unterschiedlich interpre-

²¹ Vgl. Gödert: Sprache und Bedeutung, WS 2003/04, S. 7

²² Vgl. ebd., S. 8

²³ Gödert: Der konstruktivistische Ansatz für Kommunikation und Informationsverarbeitung, 2002, S. 205

²⁴ Ebd., S. 206

²⁵ Vgl. ebd., S. 207

²⁶ Vgl. ebd., S. 211

²⁷ Beier, Heiko: Intelligente Informationsstrukturierung und TextMining mit semantischen Netzen. Beitrag zur 25. DGI-Online-Tagung Competence in Content, S. 3

²⁸ Ebd.

²⁹ Vgl. Gödert: Der konstruktivistische Ansatz für Kommunikation und Informationsverarbeitung, 2002, S. 212

³⁰ Vgl. Gödert: Der konstruktivistische Ansatz für Kommunikation und Informationsverarbeitung, 2002, S. 214

tiert werden, sondern dass darüber hinaus Begriffe aus dem semantischen Umfeld mitgedacht werden. Für eine Suche kann das bedeuten, dass der Suchbegriff mit dem semantischen Umfeld gleichgesetzt wird und auch Titel gefunden werden sollen, die durch Begriffe aus diesem Kontext beschrieben werden.³¹ Beispielsweise können zu einer Suche nach „Thesaurus“ auch Titel zur Tätigkeit „Indexierung“ gewünscht werden, ohne dass dies in der Suchanfrage explizit benannt wäre. Ebenfalls denkbar wäre auch eine sehr individuelle Beziehung von „Thesaurus“ zu „Muppet Show“, denn „welche Zusammenhänge einem Menschen wichtig erscheinen oder welche er für unbedeutend hält, ergibt sich aus der Erfahrung des Menschen – dem Kontext eingepprägter Wahrnehmungen.“³² Die Beziehungen, die zwischen Begriffen in Dokumenten bestehen, lassen sich unterscheiden in Zusammenhänge, die „an sich“ bestehen oder in solche, die aus dem vorliegenden Kontext entstanden sind. „Fest verdrahtete“³³ Begriffsrelationen wie hierarchische Zusammenhänge werden paradigmatische Beziehungen genannt. Dagegen beschreiben syntagmatische Beziehungen Begriffsrelationen, die durch den konkreten Kontext entstanden sind.³⁴ Das Syntagma kann einen semantischen Zusammenhang zwischen Begriffen erzeugen, der ebenfalls für die Suche von Interesse sein kann. Aber anders als paradigmatische Relationen sind syntagmatische Relationen dadurch, dass sie aus dem Kontext entstehen, nicht für jeden gleich nachvollziehbar.

2.2.2 Suchprozess im Information Retrieval

Zu der Schwierigkeit, dass Sprache uneindeutig ist und sowohl Fragen als auch Dokumentinhalte vielfältig beschrieben werden können, kommt im Information Retrieval das Problem der vagen Suchinteressen hinzu. In Kapitel 2.1 wurde bereits erwähnt, dass es typisch für vage Anfragen ist, dass sie nicht eindeutig definiert sind und nur im Dialog mit dem System in einem iterativen Reformulierungsprozess beantwortet werden können. Es handelt sich um Fragen, deren Beantwortungen nicht klar vorhergesagt werden können. Der Suchende weiß nicht genau, welche Antwort seinen Informationsbedarf decken könnte, denn es liegt kein konkreter Informationsbedarf im Sinne einer Faktenfrage vor, wie beispielsweise die Frage nach einer Telefonnummer bei der Auskunft, sondern ein problemorientierter Informationsbedarf, dessen thematische Grenze nicht exakt bestimmt werden kann.³⁵

Als Auslöser für einen Suchprozess kann ein Wissensdefizit vorliegen, zu dem eine Antwort gewünscht wird. Um den Suchprozess zu beginnen, wird das Suchbedürfnis als eine Frage formuliert, die vom Informationssys-

³¹ Vgl. Gödert, Winfried; Lepsky, Klaus: Semantische Umfeldsuche im Information Retrieval in Online-Katalogen, 1997, S. 7

³² Beier: Intelligente Informationsstrukturierung und TextMining mit semantischen Netzen, S. 3

³³ Stock, Wolfgang: Wissensrepräsentation. In: Lehre. Theoretische Informationswissenschaft, S. 77

³⁴ Vgl. Gödert: Sprache und Bedeutung, WS 2003/04, S. 8

³⁵ Vgl. Stock, Wolfgang: Einführung in die Informationswissenschaft. In: Lehre. Theoretische Informationswissenschaft, S. 38

tem bearbeitet werden kann. Dazu ist es nötig eine gewisse Vorstellung davon zu haben, mit welchen Begriffen das Wissensdefizit in Verbindung steht, um entsprechend den Suchprozess starten zu können. Da das Wissensdefizit häufig aber nicht exakt eingegrenzt werden kann und die Schwierigkeiten der sprachlichen Benennungen hinzukommen, liefert das Informationssystem selten gleich die vollständige und einzig richtige Antwort. Es ist vielmehr ein allmähliches Annähern an die Fragestellung durch Reformulierung erforderlich, um schließlich die gewünschte Antwort zu erhalten. Dazu muss jeweils die Ergebnismenge der Suchanfrage rezipiert und verarbeitet werden. Die Informationen werden dabei mit bereits vorhandenem Wissen in Beziehung gesetzt, woraus sich ein geänderter Informationsbedarf entwickelt, der wiederum eine modifizierte Suchanfrage nach sich ziehen kann. Dieser iterative Prozess der Frageformulierung und Ergebnisrezeption soll letztlich zur Deckung des Wissensdefizits führen – vorausgesetzt die entsprechenden Informationen sind im System gespeichert und werden vom Suchenden verstanden. Das bedeutet, dass die Informationssuche durch Interaktion mit dem Retrievalsystem erfolgt, indem die Frage präzisiert, eingengt, ausgeweitet oder eine gänzlich neue Frage formuliert wird.³⁶ In der direkten Kommunikation zwischen zwei Menschen tragen beide Partner zum Gelingen des Informationsaustausches bei, indem sie durch Rückkopplung eine Vorstellung der kognitiven Strukturen des jeweiligen Anderen entwickeln.³⁷ Im Information Retrieval muss dagegen der Dialog zwischen einem Menschen und dem Computer gelingen.

2.3 Optimierungsmöglichkeiten

Im Folgenden werden Optimierungsmöglichkeiten für die Probleme im Textretrieval vorgestellt. Neben den in Information Retrievalsystemen häufig eingesetzten Suchfunktionen sind dies automatische Indexierungsverfahren zur Verbesserung der Indexterme.

2.3.1 Suchfunktionen

Standardmäßig ist in Information Retrievalsystemen boolesches Retrieval möglich. Damit wird nach dem Prinzip des Exact Matches geprüft, ob ein gesuchter Term in der Dokumentenkollektion vorhanden ist. Ist dies der Fall, wird das entsprechende Dokument als Treffer zur Suchanfrage ausgegeben. Für Abfragen von komplexen Suchthemen können mehrere Suchwörter mit den logischen Operatoren UND, ODER und NICHT verknüpft werden. Durch den Einsatz von Abstandsoperatoren kann boolesches Retrieval präzisiert werden. So kann mit einer Proximity-Suche der maximale

³⁶ Zu diesem Abschnitt vgl. Englmeier, Kurt: Ein Computer mit einem offenen Interface für Informationsprobleme. In: Mensch und Maschine. Informationelle Schnittstellen der Kommunikation, 1992, S. 229-244, hier: S. 230f.

³⁷ Vgl. Gödert: Der konstruktivistische Ansatz für Kommunikation und Informationsverarbeitung, 2002, S. 212

Abstand zwischen zwei Suchtermen festgelegt oder mit einer Adjacency-Suche zwei unmittelbar benachbarte Terme gesucht werden. Die strenge Bedingung des Exact-Matches kann durch die Suchfunktionen Maskierung und Trunkierung gemildert werden. Damit können einzelne Buchstaben oder ganze Wortenden beziehungsweise Wortanfänge durch Variablen ersetzt werden. Zusätzlich kann mit dem Angebot einer Fuzzy-Suche vom System automatisch eine vordefinierte Anzahl an Buchstabenabweichungen zwischen Suchterm und Indexterm zugelassen werden, so dass Tippfehler und Wortvarianten nicht zu einem Mismatch führen.³⁸

2.3.2 Linguistische automatische Indexierung

Neben der klassischen intellektuellen Indexierung von Dokumenten, bei der deren Inhalte durch manuell zugeteiltem Vokabular einer Indexierungssprache repräsentiert werden, können mit automatischen Indexierungsverfahren die Indexterme aus den Dokumenten aufbereitet werden, um damit die Vergleichbarkeit von Such- und Indextermen zu erhöhen. Eines der Ziele der automatischen Indexierung ist die „grammatikalische Vereinheitlichung des Suchvokabulars“³⁹ zur Reduzierung der Wortvarianten. Für den Einsatz linguistisch basierter Ansätze wird für die deutsche Sprache ein Lexikon benötigt, indem alle sprachlichen Varianten eines Wortes verzeichnet sind.⁴⁰ Für Sprachen, in denen die Wortbildung regelmäßiger als im Deutschen erfolgt, wie beispielsweise im Englischen, können die Regeln der Sprache in einem Algorithmus hinterlegt und Terme nach diesem Schema bearbeitet werden.⁴¹ Die verschiedenen Formen eines Wortes werden mittels der Wörterbucheinträge oder des Algorithmus auf die gemeinsame Grundform reduziert und Wortableitungen substantiviert.⁴² Zusätzlich können Komposita in ihre einzelnen Wortbestandteile zerlegt werden. Da eine Dekomposition nicht immer sinnvoll ist⁴³, erfolgt die Zerlegung mit Hilfe von Lexikoneinträgen. Über eine syntaktische Analyse oder wiederum über Einträge im Wörterbuch können auch Mehrwortgruppen erkannt und als Phrase indiziert werden.⁴⁴

Neben der grammatikalischen Vereinheitlichung ist die „semantische Zusammenführung des Suchvokabulars beziehungsweise Bereitstellung alternativer Suchbegriffe (Synonyme)“⁴⁵ ein weiteres Ziel der automatischen Indexierung. Dazu werden relationierte Begriffe aus einem semantischen Netz, wie beispielsweise aus einem Thesaurus, in einem Wörterbuch erfasst,

³⁸ Zu diesem Kapitel vgl. Lepsky: Methoden und Verfahren des Information Retrieval, Kapitel 2 und 4

³⁹ Lepsky, Klaus: Im Heuhaufen suchen – und finden. Automatische Erschließung von Internetquellen: Möglichkeiten und Grenzen. In: BuB 50 (1998) 5, S.336-340, hier: S. 338

⁴⁰ Vgl. Nohr, Holger: Theorie des Information Retrieval II. Automatische Indexierung. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, 2004, S. 215-225, hier: S. 221

⁴¹ Vgl. ebd., S. 220

⁴² Vgl. Lepsky: Methoden und Verfahren des Information Retrieval, S. 33 und S. 36

⁴³ Siehe dazu auch S. 5

⁴⁴ Vgl. Nohr: Theorie des Information Retrieval II, 2004, S. 222

⁴⁵ Lepsky: Im Heuhaufen suchen – und finden, 1998, S. 338

so dass sie damit indexiert werden können. Den Dokumenten können so synonyme Indexterme und beispielsweise auch in hierarchischer Beziehung stehende Begriffe zugeteilt werden.⁴⁶ Durch entsprechende Einträge können auch Übersetzungen von Termen in verschiedene Sprachen realisiert werden.

2.3.3 Statistische Automatische Indexierung

Das Generieren von Termgewichten dient der Quantifizierung, wie wichtig ein Term eines Dokuments für die Inhaltsbeschreibung ist. Die Annahme, „die Signifikanz von Wörtern für die Bedeutung eines Textes auf statistischem Wege ermitteln zu können“⁴⁷, geht auf Luhn zurück, der den Gedanken hatte, dass nicht alle Wörter in einem Text den Inhalt gleichermaßen vertreten können und dass durch Worthäufigkeiten eine Aussage zur Bedeutsamkeit getroffen werden kann. Dass Wörter nicht zufällig in Dokumenten auftreten, was eine statistische Aussage erst ermöglicht, wurde durch Zipf belegt. Er zeigte, dass die Auftretenshäufigkeit von Wörtern in einer Dokumentensammlung multipliziert mit dem Positionswert des Häufigkeitsranges einen konstanten Wert ergibt. Daraus folgt, dass die Häufigkeitsverteilung von Wörtern einer Gesetzmäßigkeit folgt und zwischen der Auftretenshäufigkeit eines Wortes und seiner Bedeutung ein Zusammenhang besteht. Für das Retrieval führt das zur Annahme, dass „die Häufigkeit eines Wortes ... über das reine Auftreten hinaus entscheidendes Kriterium für die Relevanz eines Dokuments in Bezug zum Suchterm“⁴⁸ ist. Um die Bedeutung, die ein Term für die inhaltliche Beschreibung des Dokumentes besitzt, zu bestimmen, wird die inverse Dokumenthäufigkeit (IDF) ermittelt. Dieser Wert gibt an, wie häufig der Term je Dokument in Relation zur Textlänge vorkommt und wie viele Dokumente in der Kollektion vorliegen, die diesen Term enthalten. Es wird also ein dokumentspezifischer und ein kollektionsspezifischer Wert ermittelt. Dabei wird Termen eine höhere Signifikanz zugeschrieben, je häufiger sie in einem Dokument, aber je seltener sie in der gesamten Kollektion auftreten, denn selten vorkommende Bezeichnungen haben „einen höheren Diskriminanzeffekt als häufig auftretende“⁴⁹. Somit wird der Gewichtungswert für die einzelnen Terme aus dem dokumentspezifischen Wert dividiert durch den kollektionsspezifischen Wert ermittelt. Je höher der Gewichtungswert eines Terms ist, desto relevanter ist er für die Suchanfrage. Der Gewichtungswert kann durch weitere Kriterien, wie beispielsweise die Position des Terms im Dokument, verbessert werden, indem zum Beispiel Terme aus dem Titel von vornherein als bedeutsamer bewertet werden, als Wörter aus dem Fließtext. Um die statistischen Aussa-

⁴⁶ Dass dieses Verfahren der Indexergänzung Suchen um relevante Treffermengen erweitert, wurde in Retrievaltests nachgewiesen. Vgl. dazu beispielsweise Sachse, Elisabeth; Liebig, Martina; Gödert, Winfried: Automatische Indexierung unter Einbeziehung semantischer Relationen. Ergebnisse des Retrievaltests zum MILOS II-Projekt, 1998.

⁴⁷ Nohr: Theorie des Information Retrieval II, 2004, S. 217

⁴⁸ Lepsky: Methoden und Verfahren des Information Retrieval, S. 28

⁴⁹ Nohr: Theorie des Information Retrieval II, 2004, S. 217

gen nicht durch verschiedene Wortformen zu verfälschen, ist es sinnvoll, zuerst eine linguistische Indexierung zur Wortvereinheitlichung durchzuführen. Die Verwendung von Gewichtungswerten führt beinahe zwingend zu einem Retrieval mit Ranking der Ergebnismenge.⁵⁰ Die Treffer werden sortiert nach den Gewichten der gesuchten Indexterme ausgegeben, so dass Dokumente mit hoch gewichteten Termen in der Liste an vorderer Stelle positioniert sind.⁵¹

Um mit einer statistischen Analyse qualitative Angaben machen zu können, wird eine hinreichend große Textmenge als Basis benötigt, die möglichst inhaltlich aussagekräftig ist.⁵² In Bibliotheken mit traditionellen Titelaufnahmen ist diese erforderliche Datenmenge und inhaltliche Beschreibung nicht im ausreichenden Maß gegeben, so dass für die statistische Indexierung eine Datenanreicherung notwendig ist.⁵³

2.4 Semantische Netze

Da Sprache nur im Kontext eindeutig ist und Menschen in Begriffszusammenhängen denken, können zur Dialoggestaltung zwischen Mensch und Computer im Retrievalprozess semantische Netze eingesetzt werden. Der Nutzen der semantischen Netze im Information Retrieval wird in Kapitel 2.4.1 vorgestellt. Semantische Netze können sowohl intellektuell, wie beispielsweise Thesauri, oder aufgrund statistischer Korrelationen zwischen Wörtern in Dokumenten erstellt werden. In Kapitel 2.4.2 werden Thesauri und Methoden zur statistischen Generierung von Begriffsbeziehungen vorgestellt.

2.4.1 Semantische Netze im Information Retrieval

Semantische Netze sind Begriffe im Bedeutungszusammenhang, die durch ihren Netzcharakter „verschiedene kontext-abhängige Sichten“⁵⁴ auf Informationszusammenhänge erlauben. Durch das Einbinden eines semantischen Netzes in den Retrievalprozess ist es dem Nutzer möglich, seine individuelle Begriffsstruktur mit der im System gespeicherten Struktur abzugleichen und darüber eine Suchformulierung zu erzeugen.⁵⁵ Da mit der im System angebotenen Begriffsstruktur nicht alle individuellen Begriffsbeziehungen verschiedener Nutzer wiedergegeben werden können, muss der Nutzer sich im Netz zunächst orientieren und versuchen, das Begriffsangebot zu verstehen

⁵⁰ Nohr, Holger: Maschinelle Indexierung. Skript zur Vorlesung im Fach Inhaltliche Erschließung an der Hochschule für Bibliotheks- und Informationswesen (HBI) Stuttgart, Abschnitt: Statistische Verfahren: Der Ansatz

⁵¹ Zu diesem Abschnitt vgl. Nohr: Theorie des Information Retrieval II, 2004, S. 217f., soweit dies nicht anders gekennzeichnet ist.

⁵² Vgl. Nohr: Theorie des Information Retrieval II, 2004 S. 218

⁵³ Vgl. Lepsky, Klaus: Vom OPAC zum Hyperkatalog. Daten und Indexierung. In: Erschließen, Suchen, Finden, 1996, S. 65-74, hier: S. 66

⁵⁴ Beier, Heiko: Menschengerechte Aufbereitung von Information. Vom assoziativen Gedächtnis zum Semantischen Netz, S. 9

⁵⁵ Vgl. Gödert, Winfried: Navigation und Konzepte für ein interaktives Retrieval im OPAC. Oder: Von der Informationserschließung zur Wissenserkundung. In: Mitteilungen der VÖB 57 (2004) 1, S. 70-80, hier: S. 76

und mit seiner eigenen Struktur in Übereinstimmung zu bringen. Unterstützt werden kann dieser Vorgang durch eine leichte Bedienbarkeit und durch eine intuitiv und leicht erfassbare Darstellung. Denn auch die Präsentationsform der Begriffsstruktur hat Einfluss auf die Rezeption. So werden bei einer geeigneten grafischen Darstellung einer systematischen Struktur intuitiv Unterschiede zwischen hierarchischen Beziehungen und assoziativen Querverweisen gemacht beziehungsweise können überhaupt erst erkannt werden, wohingegen bei einer sequentiellen Anordnung derselben Struktur Hierarchien oder gar assoziative Verbindungen weniger deutlich erfasst werden.⁵⁶ Die Kommunikation zwischen dem Nutzer und dem Information Retrievalsystem soll durch die angebotenen Möglichkeiten der Interaktion gelingen. Nach Fuhr führt die Unsicherheit im Retrieval zu einer Bevorzugung der interaktiven Nutzung von Informationssystemen und er fordert dafür entsprechend optimierte Schnittstellen.⁵⁷ Als prototypisch nennt er Hypertextsysteme, mit denen sich Anfragen als Folge von Interaktionen gestalten lassen.⁵⁸ Durch den Einsatz von Hyperlinkstrukturen hat der Nutzer die Möglichkeit, entlang der begrifflichen Vorgaben zu navigieren und sich an der Netzstruktur zu orientieren.⁵⁹ Der iterative Vorgang des Suchens wird dadurch unterstützt, dass die Formulierung oder Modifizierung der Frage per Begriffsauswahl aus dem semantischen Netz erfolgt. Die Anfrage muss nicht mehr nur selbst erdacht werden, sondern kann auch durch Browsen entlang der Kontextinformation formuliert werden. Dabei können dem Suchenden auch neue Aspekte aufgezeigt werden, die er ursprünglich vielleicht nicht als sein Informationsbedürfnis erkannt hatte.⁶⁰ So kann der Nutzer seine Assoziationen nicht nur mit dem Vokabular im System abgleichen und auf eine Frageformulierung abbilden, sondern sich auch durch das Begriffsangebot zu neuen Fragestellungen inspirieren lassen. Außerdem ermöglicht eine Interaktion mit dem Netz die Klärung mehrdeutiger Begriffe einer Anfrage, indem im Dialog der gewünschte Themenzusammenhang gewählt wird. Die hier erwähnten Aspekte führen zu der Annahme, dass eine durch Visualisierung und Interaktionsmöglichkeiten optimierte Schnittstelle den Suchprozess mit dem semantischen Netz im Information Retrieval sinnvoll unterstützen kann.

Ferber verweist auf eine weitere Funktion von relationiertem Vokabular in einem Retrievalsystem. Neben dem Einsatz als Hilfsmittel für die Vokabularauswahl sowohl auf Such- als auch auf Indexierungsseite kann es zusätzlich zur automatischen Erweiterung von Suchanfragen genutzt werden.⁶¹

⁵⁶ Vgl. Gödert, Winfried: Aufbereitung und Rezeption von Information. In: Medien-Informationsmanagement. Archivarisches, dokumentarisches, betriebswirtschaftliche, rechtliche und Berufsbild-Aspekte, 2003, S. 139-152, hier: S. 151

⁵⁷ Vgl. Fuhr: Konzepte zur Gestaltung zukünftiger Information-Retrieval-Systeme, 1992, S. 70

⁵⁸ Vgl. ebd., S. 62

⁵⁹ Vgl. Beier: Menschengerechte Aufbereitung von Information, S. 11

⁶⁰ Vgl. Eibl, Maximilian: Visualisierung im Document Retrieval. Theoretische und praktische Zusammenführung von Softwareergonomie und Graphik Design, 2003, S. 119

⁶¹ Vgl. Ferber: Information Retrieval, 2003, S. 56

Das automatische Einbeziehen von Äquivalenzrelationen und das Ergänzen der Suchanfrage um Oberbegriffe (Explosion) oder um Unterbegriffe (Search Down) wird auch implizites Einbeziehen genannt.⁶² Die automatische Erweiterung von Suchanfragen um assoziative Relationen kann zu einer geringeren Präzision der Treffermenge führen. Je nachdem, wie weit gefasst verwandte Begriffe im semantischen Netz ausgewiesen sind, kann dies zu einer Ausdehnung der Suche auf Themen führen, die für den Nutzer nicht relevant sind.

2.4.2 Bereitstellung semantischer Netze

Thesauri⁶³

In einem dokumentationssprachlichen Thesaurus werden für Begriffe Benennungen festgelegt und Beziehungen zwischen diesen ausgewiesen. Die terminologische Kontrolle des Vokabulars führt zur Reduzierung der Vielfalt natürlicher Sprache, indem Synonyme erfasst und Vorzugsbenennung bestimmt werden, Homonyme in ihrer Bedeutung geklärt werden und für Begriffe möglichst eindeutige Benennungen bestimmt werden. Des Weiteren umfasst die terminologische Kontrolle den Umgang mit Komposita und ihrer Zerlegung. Die Beziehungen zwischen den Begriffen werden so umfassend ausgewiesen, „dass ein semantisches Netz über den Gesamtbereich des Thesaurus entsteht.“⁶⁴ Dabei ist eine Vielzahl an Beziehungsarten denkbar; für Thesauri beschränkt man sich neben den Äquivalenzrelationen zu meist auf hierarchische und assoziative Relationen. Hierarchische Relationen weisen Unter- und Überordnungen von Begriffen aus, die generischer oder partitiver Art sein können. Bei einer generischen Relationen besitzt der untergeordnete Begriff alle Merkmale des übergeordneten Begriffs und zusätzlich mindestens ein weiteres spezifizierendes Merkmal. Partitive Relationen beschreiben Teil-Ganzes-Beziehungen, wobei der Unterbegriff ein Bestandteil des Oberbegriffs ist. Assoziationsrelationen sind Relationen zwischen Begriffen, die weder hierarchisch noch äquivalent sind, worunter unterschiedlichste Beziehungstypen verstanden werden können. Beispielsweise können als assoziative Beziehungen Relationen wie Rohstoff – Erzeugnis bis hin zu Beziehungen im Sinne von „denke auch an“⁶⁵ bestimmt

⁶² Vgl. Nohr: Maschinelle Indexierung, Abschnitt: Maschinelle Indexierung – Retrieval

⁶³ Nicht immer werden Thesauri als semantische Netze aufgefasst. Beispielsweise weist Reimer darauf hin, dass die in Thesauri beschriebenen semantischen Relationen unabhängig von dokumentspezifischen Beziehungen sind und semantische Netze dagegen über die Beschreibung von Relationen in Thesauri hinausgehen. Nach Ferber unterscheiden sich semantische Netze von Thesauri darin, dass „aufgrund der Relationen logische Schlüsse gezogen werden“ können. Siehe Reimer, Ulrich: Wissensbasierte Verfahren der Organisation und Vermittlung von Information. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, 2004, S. 155-166, hier: S. 156f. und Ferber: Information Retrieval, 2003, S. 59. In der vorliegenden Arbeit werden semantische Netze als Begriffe im Bedeutungszusammenhang aufgefasst, worunter auch Thesauri und statistisch ermittelte Begriffskontexte fallen.

⁶⁴ Burkart, Margarete: Thesaurus. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, 2004, S. 141-154, hier: S. 147

⁶⁵ Stock: Wissensrepräsentation, S. 231

werden.⁶⁶ Im Unterschied zu hierarchischen Relationen beziehen assoziative Beziehungen auch syntagmatische Relationen mit ein. Es können jedoch nicht alle erdenklichen Verbindungen ausgewiesen werden, erst recht nicht alle individuell möglichen Assoziationen und es soll kein „unüberschaubares begriffliches Dickicht“⁶⁷ entstehen. Es ist schwierig eine Grenze festzulegen, in welchem Umfang es sinnvoll ist, assoziative Relationen auszuweisen, damit sie im Retrieval die erwünschten Querverweise zu anderen Begriffen aufzeigen und ab wann das Netz unübersichtlich wird und die Orientierung verloren geht. Je besser das semantische Netz dargeboten wird und die Orientierung darin möglich ist, desto eher können auch weit gefasste Querbeziehungen aufgezeigt werden um den Suchprozess zu unterstützen. Für den Einsatz des Thesaurus zur automatischen Indexierung, automatischen Suchanfrageerweiterung und als semantische Navigationsstruktur ist die „Vollständigkeit und Qualität der Relationierung“⁶⁸ entscheidend. Das heißt, dass Vokabular sollte möglichst lückenlos erfasst und in Beziehung gesetzt sein und keine unsauber ausgearbeiteten Hierarchieleitern enthalten.

Für Bibliotheken steht vernetztes Vokabular in Form der *Schlagwortnormdatei (SWD)* zur Verfügung. Es „ist mit einem Umfang von ca. 150.000 Sachschlagwörtern das umfangreichste terminologisch kontrollierte deutschsprachige Vokabular“⁶⁹ und der Aufbau der Relationenstruktur orientiert sich seit der dritten Auflage des zugrunde liegenden Regelwerkes *Regeln für den Schlagwortkatalog (RSWK)* an der Praxis, wie sie für Thesauri gilt. Sachschlagwörter sollen demnach durchgängig relationiert und Hierarchien „vollständig vom speziellen bis zum allgemeinen Schlagwort wiedergegeben“⁷⁰ und durch assoziative Relationen ergänzt werden, um so ein lückenloses Netz zu bilden. Im Gegensatz dazu hatte die erste Auflage der *RSWK*, die sich an der Verweisungspraxis in konventionellen Schlagwortkatalogen orientierte, noch „kein lückenloses Verweisungssystem“⁷¹ zum Ziel. Aus der Diskrepanz zwischen den Bestimmungen der verschiedenen Auflagen des Regelwerkes resultieren bis heute Lücken im Verweisungsgefüge der *SWD*. In dem Projekt „Semantische Anreicherung der Schlagwortnormdatei“⁷² wurde die gegenwärtig tatsächlich erreichte Relationenstruktur der *SWD* analysiert und folgende Lücken festgestellt: 87 % der Schlagwörter sind nicht assoziativ relationiert, 38 % der Schlagwörter sind nicht hierarchisch relationiert und 34 % der Schlagwörter sind überhaupt

⁶⁶ Zu diesem Abschnitt vgl. Burkart: Thesaurus, 2004, S. 141-154, soweit dies nicht anders gekennzeichnet ist.

⁶⁷ Gödert, Winfried: Sacherschließung. Zweck, Aufgaben, Methoden: eine Vorstellung. In: Materialien zur Inhaltserschließung, WS 2003/04, S. 5

⁶⁸ Gödert; Lepsky: Semantische Umfeldsuche im Information Retrieval in Online-Katalogen, 1997, S. 15

⁶⁹ Gödert, Winfried; Hartmann, Sarah; Hubrich, Jessica; Lepsky, Klaus; Schulenburg, Karin; Trunk, Daniela: Semantische Anreicherung der Schlagwortnormdatei. Ergebnisbericht, 2004, S. 2

⁷⁰ Regeln für den Schlagwortkatalog. RSWK, 1998, § 12,3c

⁷¹ Regeln für den Schlagwortkatalog. RSWK, 1986, § 328

⁷² Durchgeführt von den Professoren Winfried Gödert und Klaus Lepsky und den Studierenden Sarah Hartmann, Jessica Hubrich, Karin Schulenburg und Daniela Trunk 2003/04 am Institut für Informationswissenschaft der Fachhochschule Köln.

nicht relationiert.⁷³ Im Vergleich dazu weisen andere Thesauri deutlich dichter relationierte Netze aus.⁷⁴ Ferner wurden Doppelhierarchien, Verweisungen auf Synonyme, zweimaliges Vorkommen gleicher Schlagwörter in einem Datensatz in unterschiedlicher Relationenfunktion, unterschiedliche Relationen zwischen analogen Begriffen und übersprungene Hierarchien ermittelt. Für den Einsatz der *SWD* als semantisches Netz im Retrieval sind diese Lücken hinderlich.⁷⁵

Statistische Generierung von Begriffsnetzen – Clustering

Das Ermitteln von Beziehungen mit statistischen Verfahren ist eine Alternative zur intellektuellen Erstellung von Begriffsnetzen. Anhand von Worthäufigkeiten in den einzelnen Dokumenten und in der gesamten Kollektion können Beziehungen zwischen Termen erkannt und die Wörter, die in einer signifikanten Beziehung zueinander stehen als Cluster ausgegeben werden. Dabei werden nur syntagmatische Beziehungen berücksichtigt.⁷⁶ Das automatische Termclustering basiert auf der Theorie, dass je öfter zwei Terme zusammen in einem Dokument vorkommen, desto höher die Wahrscheinlichkeit ist, dass sie dem gleichen Begriffskonstrukt angehören.⁷⁷ Anhand der berechneten Gewichte aus der statistischen automatischen Indexierung lässt sich durch Vergleichen der Werte bestimmen, welche Terme als ähnlich gelten, also in Beziehung zueinander stehen. Das Prinzip der Ähnlichkeitsbestimmung basiert auf der Idee des Vektorraummodells.⁷⁸ In dem von Salton entwickelten Modell werden Terme als Dimensionen aufgefasst und Dokumente durch Vektoren beschrieben.⁷⁹ Der Vektorraum, in dem die Dokumente repräsentiert werden, weist dabei so viele Dimensionen auf, wie Indexterme vorhanden sind. Werden Anfragen ebenfalls als Vektoren dargestellt, lässt sich die Ähnlichkeit zu den Dokumentenvektoren der Kollektion anhand der räumlichen Nähe bestimmen. Anders als mit einem Exact-Match-Verfahren können so zu einer Suche auch relevante Dokumente gefunden werden, die nicht die Terme aus der Anfrage enthalten. Die Berechnung der Ähnlichkeit erfolgt durch den paarweisen Abgleich des Anfragvektors mit den Dokumentenvektoren mittels des Skalarproduktes. Demnach können die Dokumente der Kollektion auch als Matrix dargestellt werden.⁸⁰

⁷³ Vgl. Gödert; Hartmann; Hubrich; Lepsky; Schulenburg; Trunk: Semantische Anreicherung der Schlagwortnormdatei, 2004, S. 6

⁷⁴ Im Projekt wurden der Infodata Thesaurus und der Standard Thesaurus Wirtschaft zum Vergleich herangezogen und die Werte für nicht assoziativ relationierte Deskriptoren von 22 % und 16 % und für gar nicht relationierte Deskriptoren von 11 % und 0,5 % ermittelt. Vgl. Gödert; Hartmann; Hubrich; Lepsky; Schulenburg; Trunk: Semantische Anreicherung der Schlagwortnormdatei, 2004, S. 7

⁷⁵ Siehe dazu auch die Darstellung der *SWD* in Kapitel 5.4, S. 67ff.

⁷⁶ Dies kann als Grundlage für eine intellektuelle Thesauruserstellung genutzt werden. Vgl. dazu beispielsweise Stock: Wissensrepräsentation, S. 283.

⁷⁷ Vgl. Kowalski, Gerald; Maybury, Mark: Information Storage and Retrieval Systems. Theory and Implementation, 2000, S. 145

⁷⁸ Vgl. ebd., S. 146

⁷⁹ Vgl. Nohr: Theorie des Information Retrieval II, 2004, S. 218

⁸⁰ Vgl. Salton; McGill: Information Retrieval, 1987, S. 85

Anhand dieser Angaben lässt sich nicht nur berechnen, wie ähnlich Dokumente zur Anfrage sind, sondern analog dazu können auch die vergebenen Indexterme in den Dokumenten als Vektoren aufgefasst werden und die Ähnlichkeit ebenfalls durch paarweisen Vergleich bestimmt werden.⁸¹ Darüber lassen sich Cluster von ähnlichen Begriffen bilden. Ein Ausschnitt einer Term-Dokument-Matrix mit den Gewichtungswerten nach IDF kann beispielsweise wie folgt aussehen:

	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8
Dokument 1	0	4	0	0	0	2	1	3
Dokument 2	3	1	4	3	1	2	0	1
Dokument 3	3	0	0	0	3	0	3	0
Dokument 4	0	1	0	3	0	0	2	0
Dokument 5	2	2	2	3	1	4	0	2

Abb. 1 Term-Dokument-Matrix.

(Quelle: Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 146)

Durch das paarweise Vergleichen der Spalten über das Skalarprodukt werden für jeden Term Angaben erzeugt, wie ähnlich er zu den anderen Termen ist. Das Ergebnis kann als eine Term-Term-Matrix wiedergegeben werden:

	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8
Term 1	7	7	16	15	14	14	9	9
Term 2	7	17	8	12	3	18	6	17
Term 3	16	8	18	18	6	16	0	8
Term 4	15	12	18	18	6	18	6	9
Term 5	14	3	6	6	16	6	9	3
Term 6	14	18	16	18	6	16	2	16
Term 7	9	6	0	6	9	2	16	3
Term 8	7	17	8	9	3	16	3	16

Abb. 2 Term-Term-Matrix

(Quelle: Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 147)

Ab wann zwei Terme als ähnlich gelten, wird durch einen Schwellenwert festgelegt.⁸² Liegt der Wert des Skalarproduktes über dem Schwellenwert, dann gilt das Termpaar als ähnlich. Bei einem Schwellenwert von bei-

⁸¹ Vgl. ebd.

⁸² Vgl. Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 146

spielsweise 10 würde sich folgende Term-Beziehungsmatrix mit eindeutigen binären Angaben aus der obigen Term-Term-Matrix ergeben:

	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8
Term 1		0	1	1	1	1	0	0
Term 2	0		0	1	0	1	0	1
Term 3	1	0		1	0	1	0	0
Term 4	1	1	1		0	1	0	0
Term 5	1	0	0	0		0	0	0
Term 6	1	1	1	1	0		0	1
Term 7	0	0	0	0	0	0		0
Term 8	0	1	0	0	0	1	0	

Abb. 3 Term-Beziehungsmatrix

(Quelle: Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 147)

Dabei weist der Wert „1“ darauf hin, dass die Terme der entsprechenden Spalte und Zeile in einer statistisch ermittelten Beziehung zueinander stehen.⁸³ Diese Ähnlichkeitsbestimmung ist die Grundlage für die zu bildenden Begriffskluster. Für die Clusterbildung können verschiedene Verfahren angewandt werden. Mit der single-link-Methode werden alle Terme in ein Cluster zusammengefasst, die jeweils mindestens zu einem anderen Term des Clusters ähnlich sind.⁸⁴ Die Terme in dieser Begriffssammlung haben nur eine schwache Beziehung zueinander, da nicht jeder Term zu jedem anderen ähnlich ist. Nach einer anderen Methode, dem cliquen-Verfahren, werden Cluster nur aus solchen Termen gebildet, die jeweils alle untereinander ähnlich sind.⁸⁵ Dadurch entstehen Cluster, die weniger Terme enthalten als solche, die nach der single-link-Methode erstellt wurden. Dafür weisen die Terme eine stärkere Beziehung zueinander auf. Zwischen diesen beiden Extremen liegen Verfahren, wie die star-Methode und die string-Methode.⁸⁶ Nach der star-Methode wird ein Term ausgewählt und mit allen zu ihm ähnlichen Termen zu einer Klasse zusammengestellt. Ein neues Cluster wird mit dem nächsten nicht zum Start-Term ähnlichen Term eröffnet. Im Algorithmus kann festgelegt werden, ob es erlaubt sein soll ein und denselben Term verschiedenen Clustern zuzuordnen. Mit der string-Methode wird beginnend mit einem Start-Term ein Cluster mit dem nächst ähnlichen Term gebildet. Ausgehend von diesem neuen Term wird wieder-

⁸³ Vgl. ebd.

⁸⁴ Vgl. Salton; McGill: Information Retrieval, 1987, S. 86

⁸⁵ Vgl. ebd.

⁸⁶ Zu den Angaben zur star- und string-Methode vgl. Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 148f.

um der nächst ähnliche, noch nicht einem Cluster zugeordnete, Term ermittelt und in die Begriffsmenge aufgenommen. Sobald kein ähnlicher Term mehr ermittelt werden kann, ist das Cluster abgeschlossen und ein neues wird eröffnet. Die zu wählende Methode der Clusterbildung sollte in Abstimmung mit dem Schwellenwert geschehen, denn mit diesem Wert wird bereits entschieden, ab wann Terme als ähnlich zueinander gelten. Alternativ kann die Clusterbildung auch mit bereits bestehenden Begriffsklustern begonnen werden. Diese Startcluster können beispielsweise jeweils alle Indexterme eines Dokuments enthalten.⁸⁷ Durch Ähnlichkeitsberechnungen werden diese dann so modifiziert, dass Cluster ähnlicher Terme entstehen. Um den Aufwand der Berechnungen gering zu halten, werden bei diesem Ansatz Zentroide stellvertretend für das gesamte Cluster zur Ähnlichkeitsbestimmung verwendet. Der Zentroid ist ein fiktiver Term, der den Durchschnittswert der Terme des Clusters repräsentiert.⁸⁸ Es wird jeweils die Ähnlichkeit der Terme zum Zentroid berechnet und aus den ähnlichen Termen neue Cluster mit neuen Zentroiden gebildet.⁸⁹ Das Verfahren wird durchgeführt bis die Cluster stabil sind.

Analog zur Ermittlung von ähnlichen Dokumenten zu einer Suchanfrage lassen sich anstelle von Termähnlichkeiten auch Dokumentähnlichkeiten berechnen und zu Clustern zusammenstellen. Diese Cluster können als ein semantisches Netz ausgegeben werden, indem die Dokumente beispielsweise jeweils durch die Titel repräsentiert werden.⁹⁰

⁸⁷ Vgl. Salton; McGill: Information Retrieval, 1987, S. 86

⁸⁸ Vgl. Kowalski; Maybury: Information Storage and Retrieval Systems, 2000, S. 151

⁸⁹ Vgl. ebd.

⁹⁰ Solche Clusterstrukturen werden nicht nur für die Darstellung einzelner Dokumente genutzt, wie dies im Brockhaus multimedial mit dem Wissensnetz geschieht und in Kapitel 5.1, S. 33ff. vorgestellt wird. Dokumentcluster können als Themengruppen durch eine charakteristische Benennung repräsentiert und die gesamte Kollektion in eine räumliche Umgebung übertragen werden. Die Visualisierung kann beispielsweise als Sternfeldsimulation oder als topografische 3D-Karte erfolgen. Siehe dazu Däßler, Rolf; Palm, Hartmut: Virtuelle Informationsräume mit VRML. Informationen recherchieren und präsentieren in 3D, 1998, S. 79ff. Zur Nutzung der Informationsraum-Metapher mit Themenkarten siehe auch Eibl: Visualisierung im Document Retrieval, 2003, S. 71ff.

3 INTERAKTION

In diesem Kapitel wird dargestellt, welche Funktionalitäten eine Rolle bei der interaktiven Nutzung semantischer Netze im Information Retrieval spielen und es wird erläutert, wie die Navigation in der Begriffsstruktur erfolgen kann. Zunächst wird der Begriff der Mensch-Computer-Interaktion und das Interface als Schnittstelle zwischen Nutzer und System vorgestellt, bevor auf verschiedene Formen der Interaktion eingegangen wird. Danach wird Hypertext als Instrument zur Interaktion in semantischen Netzen behandelt.

3.1 Mensch-Computer-Interaktion

Der Begriff Interaktion bezeichnet nach einer allgemeinen Definition im Brockhaus „Kommunizieren“⁹¹ und „aufeinander bezogenes Handeln zwischen zwei oder mehreren Personen.“⁹² Analog dazu beschreibt die Mensch-Computer-Interaktion⁹³ die „Kommunikation zwischen Mensch und Computer“⁹⁴. Damit der Mensch mit Hilfe eines Computersystems eine Aufgabe bewältigen kann, muss er mit dem System in einen Dialog treten. Dieser Dialog setzt sich aus Interaktionen zwischen ihm und dem System zusammen.⁹⁵ Eine Interaktionsabfolge liegt beispielsweise bei der in Kapitel 2.2.2 beschriebenen Informationssuche vor. Der Suchprozess wird durch das Informationsbedürfnis des Nutzers ausgelöst und seine Absicht, Dokumente zur Beantwortung zu finden, wird in eine Aktion umgesetzt, wie etwa die Eingabe einer Suchanfrage. Preim nennt solche Aktionen, die ein Nutzer durchführt, um eine Aufgabe zu bewältigen, Interaktionsaufgaben.⁹⁶ Zur Durchführung von Interaktionsaufgaben werden vom System verschiedene Interaktionstechniken bereitgestellt, wie beispielsweise die Möglichkeit zur Texteingabe oder das Selektieren von Objekten mit der Maus.⁹⁷

Der Funktionsumfang des Systems wird dem Nutzer durch das Interface⁹⁸ zugänglich gemacht. In Anlehnung an den Design-Theoretiker Gui Bonsiepe beschreibt Thissen das Interface als eine Verbindung zwischen

⁹¹ Interaktion (Artikel). In: Brockhaus multimedial 2005 premium, 2004

⁹² Ebd.

⁹³ Synonym verwendet werden auch die englischen Bezeichnungen Human-Computer-Interaction und Human-Computer-Communication.

⁹⁴ Döring, Nicola: Computervermittelte Kommunikation, Mensch-Computer-Interaktion. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, 2004, S. 351-362, hier: S. 357

⁹⁵ Vgl. Heinecke, Andreas: Mensch-Computer-Interaktion, 2004, S. 29

⁹⁶ Vgl. Preim, Bernhard: Entwicklung interaktiver Systeme. Grundlagen, Fallbeispiele und innovative Anwendungsfelder, 1999, S. 117

⁹⁷ Vgl. ebd.

⁹⁸ Neben den in dieser Arbeit betrachteten Schnittstellen zwischen Mensch und Computer, können auch weitere Schnittstellentypen unterschieden werden, wie beispielsweise Hardware-Software-Schnittstellen. Siehe dazu zum Beispiel Halbach, Wulf: Interfaces. Medien- und kommunikationstheoretische Elemente einer Interface-Theorie, 1994, S. 169

dem Nutzer, seiner Aufgabe und dem Werkzeug, dass zur Aufgabenbewältigung herangezogen wird.⁹⁹ „Das Interface ist die Schnittstelle zwischen einem Werkzeug und dem Menschen, der dieses Werkzeug benutzt.“¹⁰⁰ Durch das Interface wird die interaktive Benutzung des Werkzeugs, wie beispielsweise die Nutzung der in dieser Arbeit beschriebenen Informationssystemen, ermöglicht, indem an der Schnittstelle Daten durch Nutzereingaben und Systemausgaben übermittelt und ausgetauscht werden können.¹⁰¹ Dafür sollte das Interface optimalerweise so gestaltet sein, dass eine intuitive und leichte Handhabung möglich ist und ein ungehinderter Zugang zu den Interaktionsfunktionen gewährleistet wird.¹⁰²

Die verschiedenen Formen der Interaktionen lassen sich unterschiedlichen Stilen zuordnen. Die Möglichkeit zur Texteingabe entspricht einem sprachbasierten Interaktionsstil, während eine Zeige- und Selektionshandlung mit der Maus grafisch orientiert ist.¹⁰³ In Analogie zu dieser Zuordnung werden diverse Interfacetypen unterschieden. Die Einteilung kann entsprechend zu den unterschiedlichen, die Interaktion ermöglichenden Eingabegeräten erfolgen. In einem Character-based Interface¹⁰⁴ erfolgen Eingaben unter Verwendung einer Tastatur, dagegen werden in einem Graphical User Interface Eingaben mit einem Zeigegerät, wie beispielsweise der Maus, gemacht. Neben diesen standardmäßig eingesetzten Schnittstellen gibt es auch Speech-based User Interfaces, bei denen die Eingaben durch natürliche Sprache in ein Mikrofon erfolgt und Mental User Interfaces, die über Sensoren Gehirnstromaktivitäten des Nutzers direkt übermitteln.¹⁰⁵ Die Reihenfolge dieser Aufzählung entspricht auch der historischen Entwicklung, wobei in vielen heutigen Systemen eine kombinierte Nutzung stattfindet. Die Kombination verschiedener Interaktionsstile bietet sich vor allem dann an, wenn unterschiedliche Aufgaben zu erledigen sind oder unterschiedliche Nutzer das System bedienen.¹⁰⁶ Im Folgenden werden sprachbasierte und grafisch orientierte Interaktionsstile kurz vorgestellt.

Die sprachbasierte Interaktion kann entweder in Form von Kommandos oder in natürlicher Sprache erfolgen. Das Verwenden von Kommandos erfordert vom Nutzer Kenntnisse der jeweilig zu verwendenden Befehlssprache. Im Gegensatz zur natürlichen Sprache, die mehrdeutig und kontextabhängig ist, sind Kommandosprachen eindeutig interpretierbar, da sie künstlich für einen bestimmten Anwendungszweck erstellt wurden. Durch den Einsatz von natürlicher Sprache für die Kommunikation zwischen Mensch und Computer, beispielsweise in Form eines Frage-Anwort-DIALOGs, wird

⁹⁹ Vgl. Thissen, Frank: Kompendium Screen-Design. Effektiv informieren und kommunizieren mit Multimedia, 2003, S. 18

¹⁰⁰ Ebd.

¹⁰¹ Vgl. Döring: Computervermittelte Kommunikation, Mensch-Computer-Interaktion, 2004, S. 357

¹⁰² Vgl. ebd., S. 358

¹⁰³ Vgl. Preim: Entwicklung interaktiver Systeme, 1999, S. 525

¹⁰⁴ Synonym verwendet wird auch die Bezeichnung Command-line User Interface.

¹⁰⁵ Zu den verschiedenen Interfacetypen vgl. Döring: Computervermittelte Kommunikation, Mensch-Computer-Interaktion, 2004, S. 358

¹⁰⁶ Vgl. Shneiderman, Ben: User Interface Design, 2002, S. 99

versucht, die Kommunikation zwischen Menschen zu imitieren, um dadurch eine leichte und intuitive Bedienung zu ermöglichen. Die Eingabe muss dabei nicht zwingend über Tastatur erfolgen, sondern kann je nach Schnittstelle auch gesprochen übertragen werden.¹⁰⁷

Der Dialog zwischen Nutzer und System mit Menüs erfolgt durch Auswahl vorgegebener Optionen. Mit Hilfe eines Zeigegeräts wie der Maus oder per Shortcuts können die vom System vorgegebenen Menüeinträge selektiert werden. Gegebenenfalls kann eine Auswahl zu weiteren Untermenüs oder weiteren Eingabe- und Auswahlfeldern wie beispielsweise Formulare führen. Damit Nutzer sich leichter an die Position von Menüpunkten erinnern können, werden inhaltlich zusammengehörende Einträge in der Regel gruppiert und hierarchisch sortiert.¹⁰⁸ Menüs können auf verschiedene Arten präsentiert werden, wie beispielsweise in Form von Pull-down-Menüs, Pop-up-Menüs und als Menüs mit Radiobuttons.¹⁰⁹ Pull-down-Menüs werden durch Auswahl eines Menüeintrags aus einer statischen Menüleiste aufgeklappt, wonach die Auswahl eines weiteren Menüeintrags möglich ist. Im Gegensatz dazu sind Pop-up-Menüs nicht ständig verfügbar, sondern erscheinen nach einem Klick, standardmäßig mit der rechten Maustaste, auf oder durch Mouse-Over über ein Objekt auf dem Bildschirm. Das daraufhin geöffnete Menü ist kontextsensitiv zum ausgewählten Gegenstand.¹¹⁰ Radiobuttons erzwingen die Auswahl genau einer Option aus mehreren angegebenen Möglichkeiten.¹¹¹ Die Buttons sind als Kreise dargestellt und eine Auswahl per Maus wird durch einen schwarzen Punkt kenntlich gemacht.

Mit der Interaktionsform der direkten Manipulation hat der Nutzer „den Eindruck .. direkt mit Objekten auf dem Bildschirm zu interagieren.“¹¹² Die Interaktion basiert auf der grafischen Repräsentation von Objekten, die durch Anklicken oder Mouse-Over über dieser Darstellung eine mit dem Objekt verknüpfte Aktion auslöst.¹¹³ Dadurch, dass Aktionen direkt durch Bewegung und Auswahl mit der Maus ausgelöst werden, können Aufgaben schnell ausgeführt und die Ergebnisse der Aktion unmittelbar betrachtet werden.¹¹⁴ Heinecke sieht die wichtigsten Eigenschaften der direkten Manipulation in der ständigen Visualisierung und der damit verbundenen permanenten Verfügbarkeit der relevanten Objekte und Aktionen, in der physischen Handlung mit dem Zeigegerät anstelle von textlichen Eingaben, die möglicherweise die Kenntnis einer Kommandosprache verlangen und in der

¹⁰⁷ Zu diesem Abschnitt vgl. Preim: Entwicklung interaktiver Systeme, 1999, S. 118

¹⁰⁸ Zu diesem Abschnitt bis hierhin vgl. Preim: Entwicklung interaktiver Systeme, 1999, S. 101

¹⁰⁹ Die Auswahl dieser drei Menüformen wurde im Hinblick auf die Anwendungsbeispiele in Kapitel 5 getroffen. Weitere Darstellungsformen werden beispielsweise von Preim beschrieben, siehe dazu Preim: Entwicklung interaktiver Systeme, 1999, S. 105ff.

¹¹⁰ Zu Pop-up-Menüs und Pull-down-Menüs vgl. Heinecke: Mensch-Computer-Interaktion, 2004, S. 156

¹¹¹ Vgl. Shneiderman: User Interface Design, 2002, S. 287

¹¹² Heinecke: Mensch-Computer-Interaktion, 2004, S. 163

¹¹³ Vgl. Shneiderman: User Interface Design, 2002, S. 98

¹¹⁴ Vgl. ebd.

direkt erfolgenden Darstellung der Resultate von Aktionen.¹¹⁵ Der Effekt der unmittelbaren Rückkopplung der Aktion wird auch als Close Feedback oder Visual Feedback bezeichnet.¹¹⁶ Zu den Vorteilen des direkt-manipulativen Interaktionsstils zählt Shneiderman, dass die grafische Gestaltung „natürlicher“¹¹⁷ wahrgenommen werden kann, als reine Textinformationen, grafische Darstellungen leichter erinnert werden können und zur Exploration ermutigen, auch weil bei der Aufgabenbewältigung wenig Fehler gemacht werden können.¹¹⁸ Im einem weit gefassten Zusammenhang mit der direkten Manipulation steht das Prinzip WYSIWYG¹¹⁹, welches einen Zugriff auf Objekte und Funktionen einer Anwendung über deren visualisierten Repräsentationen ermöglicht.¹²⁰ Diese Technik hat „entscheidend zum Erfolg der grafischen Dialoggestaltung“¹²¹ beigetragen.

3.2 Interaktion in semantischen Netzen

Durchgängig angewandt wird der Interaktionsstil der direkten Manipulation mit seinem Point-und-Klickverfahren im Hypertext.¹²² Die Hypertexttechnik ist Bestandteil vieler Informationssysteme¹²³ und mit einer grafischen Darstellung und dem Prinzip der direkten Manipulation bietet sie sich für die interaktive Nutzung semantischer Netze an. Hypertext besteht aus einzelnen, in sich geschlossenen und untereinander verknüpften Texteinheiten¹²⁴.¹²⁵ Die abgeschlossenen Einheiten werden auch Knoten und die Verbindungen, die Hyperlinks, Kanten genannt. Die Kennzeichnung von verlinkten Kanten geschieht über sogenannte Anker.¹²⁶ Dies sind die anklickbaren Elemente, durch die die Navigation entlang der Kanten erfolgt und sie können beispielsweise aus Text oder Bildern¹²⁷ bestehen.¹²⁸ Als Navigation wird die „Bewegung eines Benutzers in und zwischen Medienobjekten oder Darstellungssegmenten, um einen Gegenstand, ein bestimmtes Thema oder einen spezifischen Teil der Information zu finden“¹²⁹ verstanden. Die inter-

¹¹⁵ Vgl. Heinecke: Mensch-Computer-Interaktion, 2004, S. 163

¹¹⁶ Vgl. Preim: Entwicklung interaktiver Systeme, 1999, S. 123 und Zeidler, Alfred; Zellner, Rudolf: Software-Ergonomie. Techniken der Dialoggestaltung, 1994, S. 92

¹¹⁷ Shneiderman: User Interface Design, 2002, S. 250

¹¹⁸ Vgl. ebd., 2002, S. 97

¹¹⁹ Abkürzung für What You See Is What You Get.

¹²⁰ Ursprünglich bezog sich das Prinzip nur auf die Darstellung von Texten am Bildschirm, die einer ausgedruckten Version möglichst ähnlich sein sollte.

¹²¹ Zeidler; Zellner: Software-Ergonomie, 1994, S. 91

¹²² Vgl. Kuhlen, Rainer: Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank, 1991, S. 15

¹²³ Vgl. Hammwöhner, Rainer: Hypertext. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, 2004, S. 419-428, hier: S. 419

¹²⁴ Die einzelnen Einheiten können neben Text auch multimediale Elemente beinhalten oder nur aus ihnen bestehen. Es wird dann auch die Bezeichnung Hypermedia verwendet. Siehe dazu Hammwöhner: Hypertext, 2004, S. 419

¹²⁵ Vgl. Thissen: Kompendium Screen-Design, 2003, S. 22

¹²⁶ Vgl. ebd.

¹²⁷ Diese werden auch Hotwords beziehungsweise Hotspots genannt.

¹²⁸ Vgl. Thissen: Kompendium Screen-Design, 2003, S. 78

¹²⁹ Zitiert nach Heinecke, der die ISO-Norm 1415-1 anführt. Heinecke: Mensch-Computer-Interaktion, 2004, S. 211

aktive Nutzung des Hypertextes erfolgt also durch Auswahl der Anker mit der Maus, woraufhin der Hyperlink zur Informationseinheit verfolgt und die entsprechende Einheit aufgerufen wird.¹³⁰ Von dort aus kann erneut zu weiteren Einheiten navigiert werden, wenn von ihr weitere Hyperlinks ausgehen. Charakteristisch für Hypertext ist, dass in der Regel keine lineare Lese-folge vorgegeben ist, sondern der Nutzer sich die Informationseinheiten durch Navigation selbst zusammenstellt.¹³¹ Ein typischer Effekt, der sich daraus ergibt, ist das Browsing. Es beschreibt das Navigieren zu Informationen, die zwar in einem bestimmten Kontext stehen, die Suche erfolgt jedoch nicht zielgerichtet. Nach Kühlen lassen sich vier Arten des Browsers unterscheiden: Gerichtetes Browsing mit Mitnahmeeffekt, gerichtetes Browsing mit Serendipity-Effekt, ungerichtetes Browsing und assoziatives Browsing. Das Browsing mit Mitnahmeeffekt beschreibt das Durchstöbern einer bereits thematisch zusammengestellten Menge. Dabei wird von einer zielgerichteten Suche ausgegangen, bei der der Suchende dann auf Informationen aus dem Umfeld stößt und diese „mitnimmt“. Werden dabei zufällig Informationen gefunden, die sich als „glückliche Entdeckung“ erweisen und auf die der Suchende seine volle Aufmerksamkeit richtet und darüber das ursprüngliche Suchinteresse nicht weiter verfolgt, ist der Serendipity-Effekt eingetreten. Beim ungerichteten Browsen liegt dem Suchenden nur ein sehr unklar umrissenes Informationsbedürfnis vor, das eine schrittweise Annäherung zur Beantwortung erfordert. Dies kann zu einem assoziativen Browsing führen, bei dem sich der Suchende durch die verlinkten Angebote leiten lässt und sich dabei vom eigentlichen Suchinteresse wegbewegt. Dies kann sowohl positiv sein und Interesse an weiteren Themen wecken, als auch zu dem Gefühl führen, „lost in hyperspace“ zu sein, bei dem sich der Suchende in den Informationen „verliert“.¹³² Orientierung können zum Beispiel Dialoghistorien und Back-track-Funktionen verschaffen, mit denen der Suchende zu bereits aufgerufenen Informationseinheiten zurückkehren kann.¹³³ Orientierung kann auch die Struktur selbst bieten. Beispielsweise kann sie linear aufgebaut sein oder in Form von Baumstrukturen hierarchische Verbindungen darstellen.¹³⁴ Inhaltlich verschiedene Arten von Verknüpfungen können durch visuelle Elemente unterschieden werden.

Die Interaktion in semantischen Netzen kann durch Hypertexttechnik adäquat ermöglicht werden. Dabei dient eine textliche oder grafische Repräsentation der Begriffe als Anker, von denen aus Hyperlinks zu den Knoten führen. Die Knoten können entweder die Begriffe des Netzes selbst sein, oder auch Texteinheiten, zu denen die Begriffe führen. Das Prinzip der direkten Manipulation gestattet es dem Nutzer, sich durch Anklicken der Anker durch das Netz zu bewegen und Aktionen auszulösen. Beispielsweise

¹³⁰ Vgl. Kühlen: Hypertext, 1991, S. 15

¹³¹ Vgl. Heinecke: Mensch-Computer-Interaktion, 2004, S. 25

¹³² Zu den verschiedenen Arten des Browsers vgl. Kühlen: Hypertext, 1991, S. 128ff.

¹³³ Vgl. Hammwöhner: Hypertext, 2004, S. 426

¹³⁴ Vgl. Heinecke: Mensch-Computer-Interaktion, 2004, S. 214

kann durch Selektion eines Begriffs das Begriffsumfeld dazu aufgebaut werden oder es kann der Begriff für die Formulierung einer Suchanfrage übernommen werden. Ein semantisches Netz mit Hypertexttechnik ermöglicht durch Navigation und Browsing ein exploratives und assoziatives Suchen.

4 VISUALISIERUNG

In diesem Kapitel geht es um Visualisierung als Mittel zur Unterstützung kognitiver Fähigkeiten. Der Bereich der Informationsvisualisierung, zu dem Visualisierungen in Information Retrievalsystemen zählen, wird vorgestellt und anschließend auf Visualisierungsmöglichkeiten semantischer Netze und auf Probleme bei der Darstellung eingegangen.

4.1 Informationsvisualisierung

Visualisierung ist keine Erfindung der heutigen Zeit. Menschen visualisieren Dinge aus ihrer Umwelt bereits seit Tausenden von Jahren, wovon Höhlenzeichnungen der Urmenschen zeugen. Sie können als frühe Darstellungsform visueller Kommunikation aufgefasst werden.¹³⁵ Nach Däblier ist Visualisierung zu verstehen als „Ausdruck des menschlichen Vermögens ca. 75% der Information aus visuellen Eindrücken der Realwelt aufzunehmen“¹³⁶ und es ist somit naheliegend, dass der Mensch Informationen aus seinem Erfahrungsbereich in visuelle Darstellungsformen transformiert. Visualisierung kann als ein externes Hilfsmittel die kognitiven Fähigkeiten unterstützen beziehungsweise bestimmte Erkenntnisse überhaupt erst ermöglichen.¹³⁷ So können grafische Darstellungen Ideen und Vorstellungen zum einen nach dem Motto „Ein Bild sagt mehr als tausend Worte“ vermitteln und zum anderen helfen, Vorstellungen und Ideen zu entwickeln.¹³⁸ Wie wichtig externe Mittel zur Unterstützung der Wahrnehmung und des Denkens sind und wie sehr interne und externe Repräsentationen mit den kognitiven Prozessen verwoben sind, veranschaulichen Card u.a. am Beispiel einer Multiplikationsaufgabe. Eine einfache Multiplikation kann unter Zuhilfenahme von Papier und Bleistift sehr viel schneller als durch Kopfrechnen gelöst werden und eine anspruchsvollere Aufgabe kann ohne externe Unterstützung überhaupt nicht berechnet werden. Dabei ist die Multiplikation selbst nicht das Schwierige, sondern das Merken, wieder Abrufen und Einsetzen von Zwischenergebnissen. Durch die Auslagerung der Ergebnisse auf Papier kann das Gedächtnis unterstützt werden und so zur Lösung der Aufgabe beitragen.¹³⁹ Den Nutzen von Visualisierung fasst Eibl wie folgt zusammen: „Die menschliche Kognition ist folglich auf externe Repräsentationsformen angewiesen und aufgrund der visuellen Fähigkeiten des Gehirns bietet sich

¹³⁵ Vgl. Däblier, Rolf: Informationsvisualisierung. Stand, Kritik und Perspektiven, S. 2

¹³⁶ Ebd.

¹³⁷ Vgl. Card, Stuart; Mackinlay, Jock; Shneiderman, Ben (Verf. und Hrsg.): Readings in Information Visualization. Using Vision to Think, 1999, S. 1

¹³⁸ Vgl. ebd.

¹³⁹ Zu diesem Beispiel vgl. Card; Mackinlay; Shneiderman (Verf. und Hrsg.): Readings in Information Visualization, 1999, S. 1f.

eine graphische Repräsentation an.“¹⁴⁰ Um auf die Bedeutung visueller Metaphern für die Rezeption von Informationen zu verweisen, stellt Däßler eine alte grafische Darstellung des Russlandfeldzuges von Napoleon vor, in der von mehreren Faktoren abhängige Angaben verzeichnet sind. Nur in dieser visuellen Präsentation ist es möglich, die Aspekte und Zusammenhänge der Informationen unmittelbar zu erfassen.¹⁴¹ Durch grafisch dargestellte Daten können zum Beispiel Muster wie Häufungen und Gruppierungen schneller erkannt werden, Trends sichtbar gemacht werden und Beziehungen zwischen Daten aufgezeigt werden.¹⁴² Das Erfassen von Informationen wird durch Visualisierung erleichtert, indem „der *bewußte* Denkvorgang des Benutzers in einen intuitiven Wahrnehmungsvorgang überführt wird.“¹⁴³ Visualisierung kann sich somit positiv auf die Interpretation von Informationen auswirken und dadurch zu einer optimierten Gestaltung eines Interfaces beitragen.

Bei der computerbasierten Visualisierung haben sich zwei Bereiche herausgebildet. Zunächst entwickelte sich in den 1980er Jahren die wissenschaftliche Visualisierung physikalischer Daten.¹⁴⁴ Anfang der 1990er Jahre etablierte sich dann die Informationsvisualisierung zur Darstellung abstrakter Daten, worunter auch die Daten aus Information Retrievalsystemen fallen.¹⁴⁵ Card u.a. definieren die Informationsvisualisierung als „the use of computer-supported, interactive, visual representations of abstract data to amplify cognition“¹⁴⁶ und verweisen auch auf den Aspekt der Interaktion im Zusammenhang mit computerbasierter Visualisierung. Die Repräsentation von abstrakten Daten, wie Dokumente und Begriffe aus Informationssystemen ist schwierig, denn anders als bei wissenschaftlicher Visualisierung, deren Daten sich meist räumlich in Koordinatensystemen darstellen lassen, drängt sich nicht eine bestimmte Präsentationsform auf.¹⁴⁷ Wie die Informationen dargestellt werden hängt auch vom zugrunde liegenden Datenmodell ab.¹⁴⁸ Die für die Informationsvisualisierung wichtigen Datenstrukturen hat Shneiderman identifiziert und zusammengestellt. Er führt sieben Datentypen auf, in denen Informationen durch bestimmte Attribute beschrieben werden und auf die im Retrieval zugegriffen wird. Seine Taxonomie stellt dabei ein Modell dar und verschiedene Datentypen können in konkreten Anwendungen kombiniert oder in Variationen benutzt werden.¹⁴⁹ Shneiderman unter-

¹⁴⁰ Eibl: Visualisierung im Document Retrieval, 2003, S. 44f.

¹⁴¹ Vgl. Däßler: Informationsvisualisierung, S. 2

¹⁴² Vgl. Eibl: Visualisierung im Document Retrieval, 2003, S. 43

¹⁴³ Schwartz, Dieter: Mehr Information durch Visualisierung von Daten? Konventionelle und innovative Visualisierungstechniken. In: B.I.T.online 2 (1999) 4, S. 455-462, hier: S. 455

¹⁴⁴ Vgl. Däßler; Palm: Virtuelle Informationsräume mit VRML, 1998, S. 40

¹⁴⁵ Vgl. ebd.

¹⁴⁶ Card; Mackinlay; Shneiderman (Verf. und Hrsg.): Readings in Information Visualization, 1999, S. 7

¹⁴⁷ Vgl. Eibl: Visualisierung im Document Retrieval, 2003, S. 45f.

¹⁴⁸ Vgl. Däßler: Informationsvisualisierung, S. 41

¹⁴⁹ In einem studentischen Projekt wurde, unter Ben Shneidermans Leitung an der Universität Maryland, Fachbereich Computer Science, eine Zusammenstellung an Forschungsprojekten, kommerzieller Anwendungen und wichtigen Veröffentlichungen zur Informationsvisualisierung geordnet nach den hier genannten Datentypen gemacht. Siehe dazu „Homepage OLIVE“

scheidet eindimensionale, zweidimensionale, dreidimensionale, zeitliche, multidimensionale, Verzweigungsbaum- und Netzwerkdaten. Eindimensionale Daten sind linear organisiert und können zeilenweise dargestellt werden, wie beispielsweise Text. Zweidimensionale Daten lassen sich in einer Ebene anordnen, wie beispielsweise geographische Daten. Für das Information Retrieval könnten zum Beispiel Dokument-Begriffsrelationen zweidimensional visualisiert werden. Dreidimensionale Daten beschreiben räumliche Objekte aus der Umwelt und werden häufig in der wissenschaftlichen Visualisierung eingesetzt. Es lassen sich beispielsweise aber auch abstrakte Daten wie hierarchische Begriffsstrukturen dreidimensional präsentieren. Zeitliche Daten sind eindimensionale Daten, die zusätzlich einen Zeitraum beschreiben. Solche Daten kommen beispielsweise im Projektmanagement vor. Multidimensionale Daten weisen mehr als drei Attribute auf. Die zusätzlichen Dimensionen lassen sich zum Beispiel durch Kodierungen wie Farbe und Form ausdrücken. Verzweigungsbaumdaten sind hierarchisch angeordnet, wie beispielsweise Klassifikationsstrukturen. Jedes Element der Sammlung, außer der Wurzel, hat eine Verbindung zu einem übergeordneten Element. Netzwerkdaten sind Daten, bei denen die Beziehungen zwischen Elementen nicht durch eine Verzweigungsbaumstruktur beschrieben werden können, sondern in beliebiger Anzahl untereinander verbunden sind, wie beispielsweise Begriffe in semantischen Netzen.^{150 151}

Die Wahl einer adäquaten Form der Visualisierung zum Anwendungsproblem ist entscheidend, da nicht jede grafische Umsetzung per se gelungen ist und den Interpretationsprozess unterstützt. Wie wichtig die Wahl der richtigen Präsentation sein kann, verdeutlichen Card u.a. anhand des missglückten Starts der Challenger¹⁵². Es stellte sich damals die Frage, ob der Shuttle trotz des vorherrschenden kalten Wetters starten sollte. Man befürchtete, dass die kühle Temperatur zur Beschädigung der Dichtungsringe an den Startraketen führen könnte. Als Grundlage für die Entscheidung, ob ein Risiko besteht oder nicht, wurde eine Darstellung verwendet, die die Außentemperatur und die beschädigte Dichtungsringen in den Raketen vergangener Starts auflistete. Dabei waren die für diese Entscheidung wichtigen Daten Temperatur und Grad der Beschädigung nicht grafisch umgesetzt, sondern als Ziffern angegeben. Durch die nicht vorhandene Visualisierung der entscheidenden Daten ließen sich keine Muster erkennen und es wurde in der Darstellung der Eindruck erweckt, als würden Schäden nur selten auftreten. Wären dagegen die wichtigen Variablen in der richtigen

¹⁵⁰ Zur aufgeführten Zusammenstellung der Datentypen vgl. Shneiderman: User Interface Design, 2002, S. 596ff.

¹⁵¹ Zusammenstellungen von verschiedenen Praxisbeispielen aus dem Bereich der Informationsvisualisierung werden beispielsweise dokumentiert in: Bederson, Benjamin; Shneiderman, Ben (Verf. und Hrsg.): The Craft of Information Visualization. Readings and Reflections, 2003 ; Card; Mackinlay; Shneiderman (Verf. und Hrsg.): Readings in Information Visualization, 1999 ; Däßler; Palm: Virtuelle Informationsräume mit VRML, 1998 und Eibl: Visualisierung im Document Retrieval, 2003

¹⁵² Spaceshuttle der NASA, 1986 beim Start explodiert.

Form präsentiert worden, hätte das Risiko erkannt werden können und die Freigabe zum Start wäre unwahrscheinlich gewesen.¹⁵³

4.2 Visualisierung semantischer Netze

Begriffsbeziehungen, wie sie in semantischen Netzen vorliegen, können aus einer Kombination verbaler Ausdrücke und visualisierter Struktur des Beziehungsgeflechts dargestellt werden. Die Struktur lässt sich durch eine Graphendarstellung wiedergeben und ist in einer solchen Form direkt lesbar. Graphen bestehen aus Punkten¹⁵⁴ zur Darstellung der Begriffe und aus Verbindungslinien¹⁵⁵ zur Darstellung der Beziehungen.¹⁵⁶ Unterschiedliche Relationenarten können durch grafische Mittel wie Farben und Muster kodiert oder durch textliche Beschriftung kenntlich gemacht werden. Anstelle der Linien können auch die Punkte und Begriffe selbst unterschiedlich koloriert werden. Unterschiedliche Gewichtungen der Beziehungen können beispielsweise durch unterschiedliche Linienstärken oder durch unterschiedliche Schriftgrößen der Bezeichnungen repräsentiert werden. Hierarchische Strukturen, wie sie in Thesauri mit Ober- und Unterbegriffsrelationen vorliegen, lassen sich als Verzweigungsbäume nach Art von Dateimanagern darstellen.¹⁵⁷ Auf diese Weise werden hierarchisch relationierte Begriffe beispielsweise mit der Thesaurussoftware *MIDOSThesaurus* der Firma *PROGRIS* visualisiert. Es ist über diese Anzeige „auf den ersten Blick erkennbar, auf welcher Ebene der Term in der Hierarchieleiter steht, d.h. wie er in eine Struktur von Ober- und Unterbegriffen eingebettet ist“¹⁵⁸, wie nachfolgend zu sehen ist:

¹⁵³ Dieses Beispiel ist zitiert nach Card u.a., die eine Beschreibung von Tufte aufgreifen. Vgl. Card; Mackinlay; Shneiderman (Verf. und Hrsg.): *Readings in Information Visualization*, 1999, S. 4f.

¹⁵⁴ Die Punkte werden analog zum Hypertext auch Knoten genannt.

¹⁵⁵ Die Verbindungslinien werden analog zum Hypertext auch Kanten genannt.

¹⁵⁶ Vgl. Graph (Artikel). In: *Brockhaus multimedial 2005 premium*, 2004

¹⁵⁷ Vgl. Shneiderman: *User Interface Design*, 2002, S. 607

¹⁵⁸ Stys, Thomas; Klos, Johanna; Kunkel, Paul: *Thesauri entwickeln und anwenden mit MIDOSThesaurus*. In: *Information - Wissenschaft und Praxis* 54 (2003) 5, S. 273-280, hier: S. 277



Abb. 4 Baumdarstellung in MIDOSThesaurus¹⁵⁹

An der obigen Abbildung zeigt sich jedoch auch das Problem bei der Darstellung vieler Begriffe auf einer begrenzten Bildschirmfläche. Begriffssammlungen wie Thesauri weisen in der Regel flache aber breite Hierarchien mit vielen Begriffen auf einer Ebene nach, die aufgrund des Platzmangels nicht alle gleichzeitig angezeigt werden können. Um unhandliches Scrollen zu vermeiden und trotzdem alle Begriffe darstellen zu können, gibt es Versuche, die Informationen platzsparend zu präsentieren. In Anlehnung

¹⁵⁹ Alle Abbildungen und Screenshots der Anwendungsbeispiele in Kapitel 4 und 5 sind als jpg-Dateien im Anhang hinterlegt. Vgl. „Abbildungen“

an die Fischaugentechnik aus der Fotografie wird die Begriffsstruktur so dargestellt, dass die Bestandteile, auf die sich der Nutzer konzentriert, fokussiert und die restlichen Informationen des Kontextes verzerrt dargestellt werden.¹⁶⁰ Durch Interaktion können die Begriffe aus dem Kontext in den Fokus gesetzt werden. Ein Beispiel für die verzerrte Visualisierung hierarchischer Strukturen sind hyperbolische Baumdarstellungen, bei der die Informationen auf die Oberfläche einer Kugel projiziert werden.¹⁶¹ Die nachfolgende Abbildung zeigt eine hierarchische Begriffsstruktur in einer dreidimensionalen hyperbolischen Darstellung mit dem Tool *H3*:

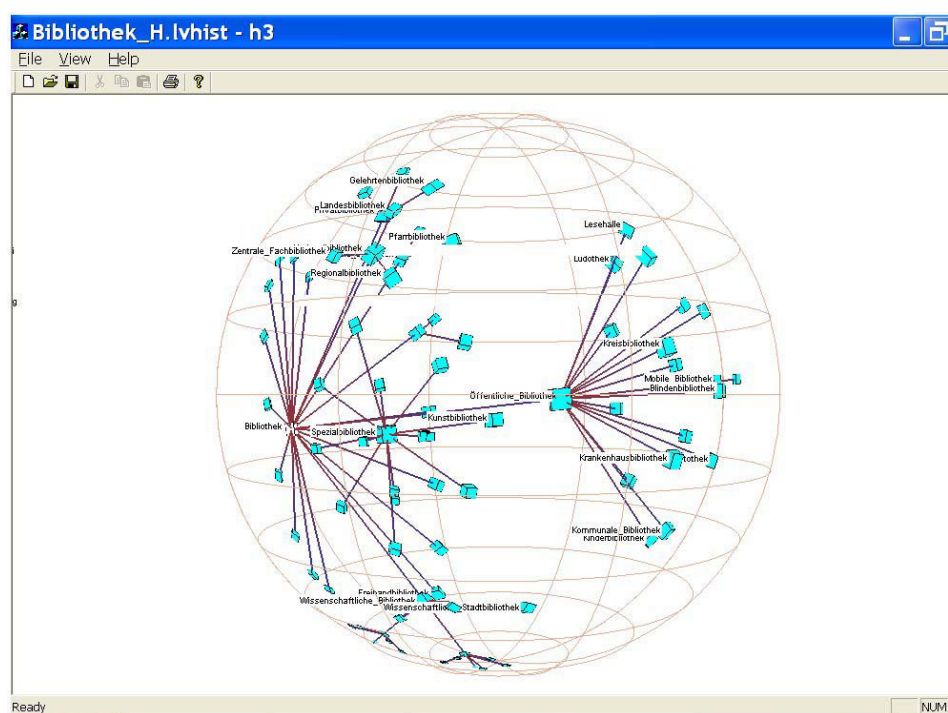


Abb. 5 Baumdarstellung in H3

somit nachvollziehbar¹⁶² erfolgt, wird Animation eingesetzt. Ein fließender Übergang von der alten zu neuen Anzeige ermöglicht dem Betrachter, den Vorgang zu beobachten und, anders als bei einem abrupten Wechsel, zu verfolgen und zu erkennen, was geschehen ist.¹⁶³ Auch bei nicht verzerrten Darstellungen kann Interaktion als Mittel zur Änderung der Perspektive auf das Vokabular eingesetzt werden. Beispielsweise können auf diese Weise sich überlappende Begriffe lesbar gemacht werden.

¹⁶⁰ Vgl. Preim: Entwicklung interaktiver Systeme, 1999, S. 361

¹⁶¹ Vgl. Däßler: Informationsvisualisierung, S. 8

¹⁶² Eibl: Visualisierung im Document Retrieval, 2003, S. 54

¹⁶³ Vgl. Preim: Entwicklung interaktiver Systeme, 1999, S. 373

Neben hierarchisch relationierten Begriffen liegen in semantischen Netzen weitere Beziehungsarten vor, wie beispielsweise in Thesauri die verwandten Begriffe und Synonyme. Diese lassen sich nicht in die Baumstruktur integrieren, da sie keine Ober- oder Unterordnung darstellen. Des Weiteren zeichnen sich Netzstrukturen dadurch aus, dass die Beziehungen zwischen den Elementen nicht durch eine Baumstruktur dargestellt werden können, da diese in beliebiger Anzahl untereinander verknüpft sein können.¹⁶⁴ Die Darstellung unterschiedlicher Relationentypen kann durch die bereits erwähnten Linien, Visualisierungselemente wie Farbe oder Muster und Beschriftung erfolgen. Die Visualisierung von Begriffen unterschiedlicher Beziehungsarten zu einem Term ist jedoch schwierig und erzeugt schnell eine unübersichtliche Darstellung, die nicht mehr erfassbar ist. Die folgende Abbildung veranschaulicht dieses Problem anhand eines Vernetzungsdiagramms für Webseiten mit einigen Knoten und vielen Verbindungslinien:

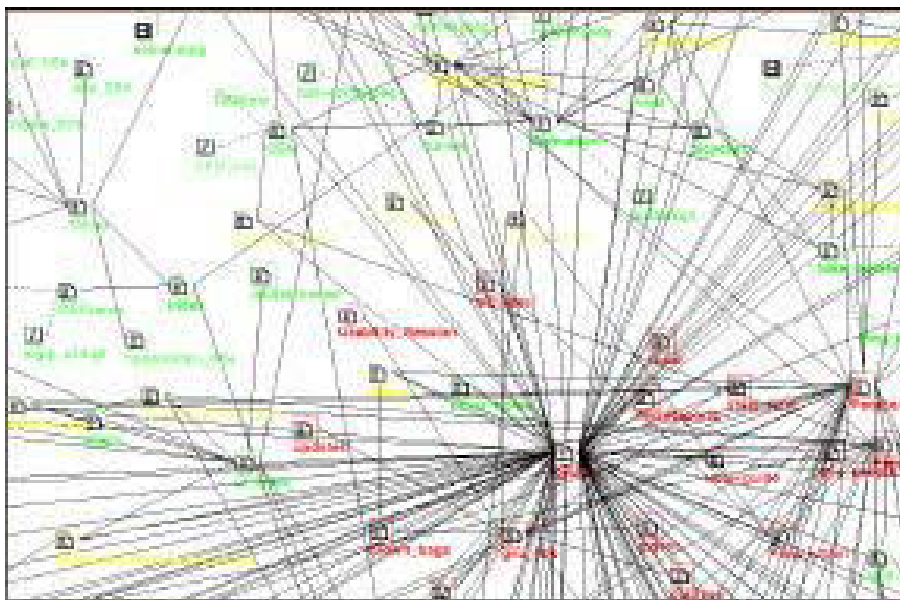


Abb. 6 Unübersichtliche Netzdarstellung mit vielen Knoten und Linien
(Quelle: Däßler: Informationsvisualisierung, 1999, S. 16)

Sollen zusätzlich Angaben zur Art der Verbindungslinien visualisiert werden, gerät die Darstellung chaotisch. Um dies zu vermeiden wird sich in konkreten Anwendungen häufig damit beholfen, die Begriffsmenge strikt auf eine geringe Anzahl festzulegen und somit das Begriffsumfeld im Hinblick auf eine bessere Darstellbarkeit zu vereinfachen. Darüber hinaus ist auch zu beachten, dass eine Visualisierung nur so gut sein kann, wie es die Datengrundlage zulässt. Sind Mängel in der Struktur vorhanden, werden

¹⁶⁴ Vgl. Shneiderman: User Interface Design, 2002, S. 609

diese durch die bessere Interpretierbarkeit von grafischen Darstellungen geradezu aufgedeckt.¹⁶⁵

¹⁶⁵ Vgl. Däbler: Informationsvisualisierung, S. 14

5 ANWENDUNGSBEISPIELE

In diesem Kapitel werden vier Anwendungsbeispiele vorgestellt, die grafisch visualisierte semantische Netze zur Unterstützung des Suchprozesses anbieten. Alle Anwendungen sind interaktiv nach dem Prinzip Hypertext zu bedienen. Die Beispiele unterscheiden sich nach Art der Netzgenerierung, nach Darstellung der Begriffsstruktur und nach vorhandenem Interaktionsumfang.

5.1 Wissensnetz im Brockhaus multimedial

Die Enzyklopädie *Brockhaus multimedial* des Verlages *Bibliographisches Institut & F.A. Brockhaus* bietet seit der Version 2001¹⁶⁶ ein Wissensnetz an, das thematisch verwandte Begriffe und Themen zu einer Suchanfrage grafisch visualisiert anzeigt und für eine Navigation durch die Artikel genutzt werden kann. In der Version 2005 sind im *Brockhaus multimedial* 240.000 Artikel¹⁶⁷ enthalten, die alle über das Netz in Beziehung zueinander stehen¹⁶⁸.

Die Begriffe im Netz werden durch Dokumentclustering der Artikel erzeugt. Nach einer linguistischen Indexierung der Artikel werden die Terme mit IDF und unter Einbezug weiterer Relevanzfaktoren, wie beispielsweise der Häufigkeitsverteilung des Wortes innerhalb eines Artikels oder von Klassifikationsinformationen, gewichtet. Anhand dieser Daten wird über Dokumentclustering ermittelt, welche Artikel ähnlich zueinander sind. Es liegen dann zu jedem Artikel Angaben vor, welche anderen Artikel zu dessen thematischem Umfeld gehören und möglicherweise für eine Suchanfrage relevant sind. Die 8 beziehungsweise 24 höchst gewichteten Artikel werden für das Wissensnetz verwendet und über deren Lemmata ausgegeben. Zum Zeitpunkt der Suche wird also auf eine bereits fertig gestellte Liste von verwandten Artikeln zugegriffen. Das Wissensnetz zu einem aufgerufenen Artikel ist bei jedem erneuten Aufrufen identisch und die Anzeige ist auf ein generiertes Netz beschränkt, das heißt, es können nicht zwei Netze miteinander kombiniert angezeigt werden.¹⁶⁹

¹⁶⁶ Vgl. „Homepage des Verbandes Bildung und Erziehung Landesverband NRW – Neuheiten des Brockhaus multimedial 2001“

¹⁶⁷ Vgl. „Homepage des Brockhaus multimedial – Produktdaten“

¹⁶⁸ Vgl. „Hilfetext im Brockhaus multimedial 2005 premium – Wissensnetz“

¹⁶⁹ Die Angaben zur Generierung des Wissensnetzes im Brockhaus multimedial basieren auf schriftlichen Auskünften von Herrn Arnd Frederichs (technischer Projektleiter des Brockhaus multimedial). Vgl. dazu „E-Mail Brockhaus“

Das Interface des *Brockhaus multimedial* ist in drei Flächen unterteilt.¹⁷⁰ Im mittleren Teil wird der Artikeltext und das Wissensnetz angezeigt. Auf der linken Seite befinden sich die Eingabezeile, die Liste der Treffermenge, sortiert nach Relevanz, und das Angebot an multimedialen Elementen zum aufgerufenen Artikel. Rechts sind Verweise auf weitere Brockhaus-Artikel, Webseiten, Literaturangaben und auf weitere derartige Informationen zum Thema zu finden.

Durch Eingabe eines Suchbegriffs in die Eingabezeile wird das Wort automatisch beim Eintippen zum vollständigen Eintrag ergänzt. Durch Drücken des Buttons „GO!“ wird der zugehörige Artikel und das Wissensnetz aufgerufen. Begriffe können durch boolesche Operatoren kombiniert und unter Verwendung von Abstandsoperatoren gesucht werden. Außerdem ist eine Trunkierung oder Maskierung möglich.

Beispielsweise führt eine Suche nach „Buch“ zu einer Trefferliste, wie sie in der linken Spalte des folgenden Beispiels zu sehen ist. Der erste Artikel der Ergebnismenge wird automatisch angezeigt und das Wissensnetz zu diesem Artikel ausgegeben:

The screenshot displays the 'DER BROCKHAUS multimedial 2005' interface. The top navigation bar includes 'LEXIKON', 'THEMEN', 'ZEITLEISTE', 'ATLAS', 'EXTRAS', 'LERNEN', 'PERSONLICH', and 'INTERNET'. The main search area on the left shows a search for 'Buch' with a 'GO!' button. Below the search bar, a list of search results is visible, sorted by relevance. The central part of the screen displays the article 'Buch' with a detailed definition and a 'COMPUTERGENERIERTER KONTEXT' (Computer-generated context) diagram. The diagram shows 'Buch in einem Umschlag o...' at the center, connected to various related terms like 'Bibel', 'Taschenbuch', 'Gutenberg: Der Buchdr...', 'Textilien: Geschichte de...', 'Buchdruck: Bewegliche Lettern...', 'Papyrus', 'Buchdruck: Eine Technikv...', and 'Handschrift: Buchkunde'. The right sidebar contains additional information, including 'KONTEXT', 'PREMIUMARTIKEL', 'QUELLENTEXTE', 'HÄTTEN SIE'S GEWUSST?', 'Brockhaus 1906', 'LITERATUR', 'WEB-LINKS', and 'NOTIZEN'.

Abb. 7 Brockhaus – Anzeige des Artikels „Buch“

¹⁷⁰ Diese und folgende Angaben zum Brockhaus multimedial beruhen auf Erkundungen durch die Verfasserin, soweit dies nicht anders gekennzeichnet ist. Die Angaben und Screenshots stützen sich auf Brockhaus multimedial 2005 premium. Siehe dazu Brockhaus multimedial 2005 premium, 2004

Eine Suche mit Homonymen wird im Dialog mit dem Nutzer geklärt. Durch Auswahl per Mausklick kann ein anderer Artikel aus der Trefferliste ausgewählt werden, wie zum Beispiel hier ein Artikel zu „Hans Christoph Buch“, der ebenfalls als 100%ig relevant zur Suchanfrage eingestuft wurde. Durch Anfahren eines Eintrags in der Trefferliste mit der Maus erscheint ein Pop-up-Fenster mit Informationen aus dem entsprechenden Artikel, wie beispielsweise hier zum Eintrag „Buch“:

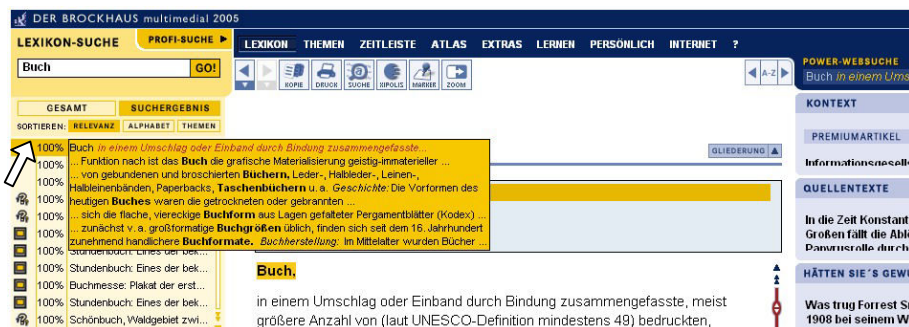


Abb. 8 Brockhaus – Pop-up-Fenster zum Eintrag „Buch“

In der Standardeinstellung des Wissensnetzes werden 8 Artikel angezeigt. Über den Button „Erweitern“ wird das Netz von 8 auf 24 verwandte Artikel erweitert. Visualisiert werden jeweils die Lemmata in sternenförmiger Anordnung um das Lemma des aktuell aufgerufenen Artikels herum, das im Zentrum des Netzes steht. Die Verbindungslinien verdeutlichen dem Nutzer, dass die Artikel aus dem Netz im Zusammenhang mit dem aufgerufenen Artikel stehen. Hier die Anzeige der Standarddarstellung des Wissensnetzes zum Artikel „Buch“:

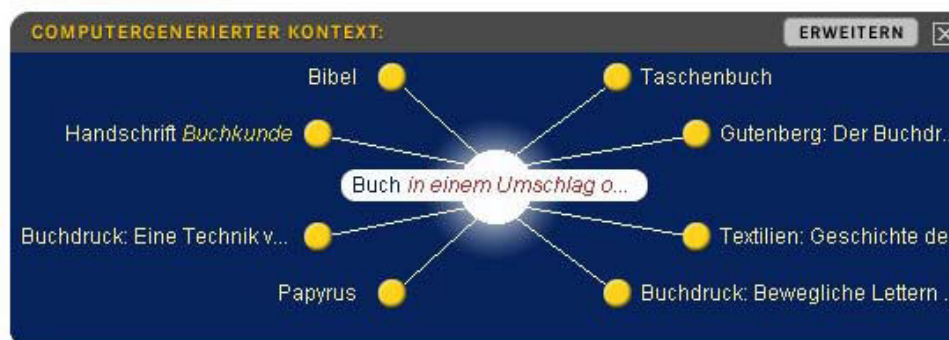


Abb. 9 Brockhaus – Wissensnetz (Standardeinstellung) zum Eintrag „Buch“

Die Anzeige der erweiterten Einstellung des Wissensnetzes zum gleichen Artikel:

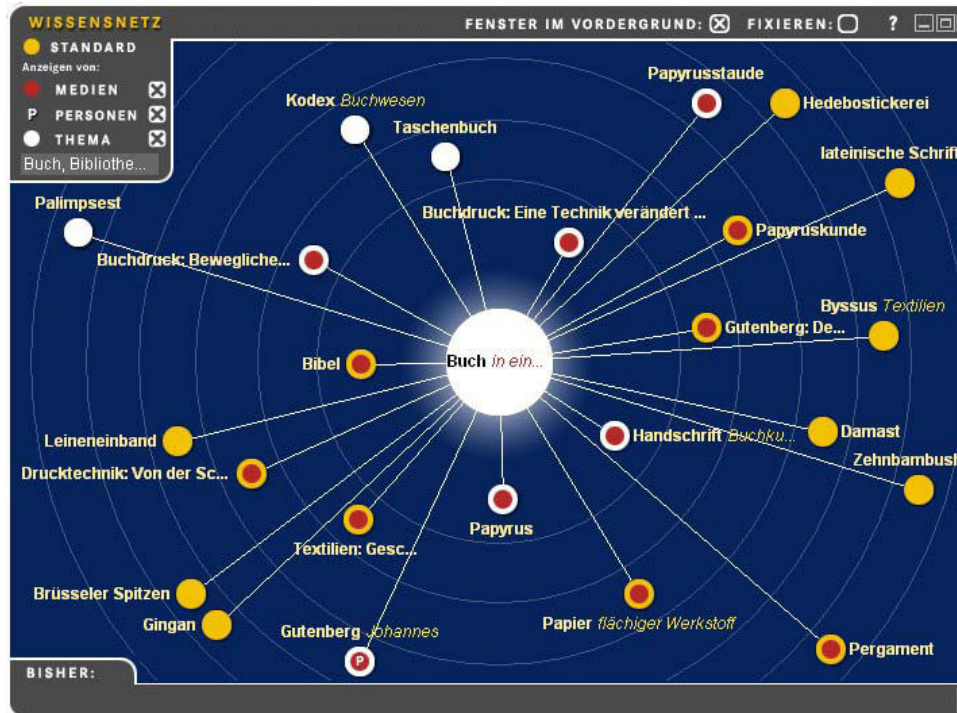


Abb. 10 Brockhaus – Wissensnetz (erweiterte Einstellung) zum Eintrag „Buch“

Alle Begriffsangaben im Netz sind zu den Artikeln verlinkt, so dass über Anklicken eines Lemmas der Artikel aus dem Netz heraus aufgerufen werden kann. Durch Anklicken eines Begriffs wird das Netz um den ausgewählten Term herum neu aufgebaut und der zugehörige Artikel aufgerufen. Im erweiterten Netz werden im Vergleich zur Standardeinstellung weitere Angaben zu den verlinkten Artikeln gemacht. Dies erfolgt visualisiert durch unterschiedliche Farbgebung der Kreise an den Begriffen oder durch den Buchstaben „P“, der biografische Artikel kennzeichnet. Die Legende, die in der linken oberen Ecke des Netzes zu sehen ist, erläutert die Bedeutung der verwendeten Farben. Bei gelben Kreisen handelt es sich um Standard-Artikel, rote Kreise verweisen auf multimediale Elemente im Artikel (wie beispielsweise Bilder, Videos oder gesprochene Texte) und die Farbe Weiß deutet auf thematisch verwandte Artikel¹⁷¹ hin. Am unteren Rand des erwei-

¹⁷¹ Alle Artikel im Brockhaus multimedial sind klassifiziert. In diesem Beispiel gelten alle Artikel als thematisch verwandt, die wie der aufgerufene Eintrag „Buch“ dem Thema „Buch, Bibliothekswesen, Schrift, Dokumentation“ zugeordnet sind. Die Klassifikationsinformationen fließen auch in die Relevanzbestimmung für die Wissensnetzgenerierung ein.

terten Netzes ist eine Leiste platziert, in der die bisher aufgerufenen Begriffe im Sinne einer Dialoghistorie aufgelistet werden.

Durch Anklicken können die schon einmal aufgerufenen Netze und Artikel erneut angezeigt werden. Durch Mouse-Over der Kreise wird eine Vorschau auf das Begriffsnetz um dieses Lemma aufgerufen und die ersten Sätze des Artikels in einem Pop-up-Fenster angegeben, wie beispielsweise im Folgenden die Vorschau auf den Eintrag „Handschrift“:

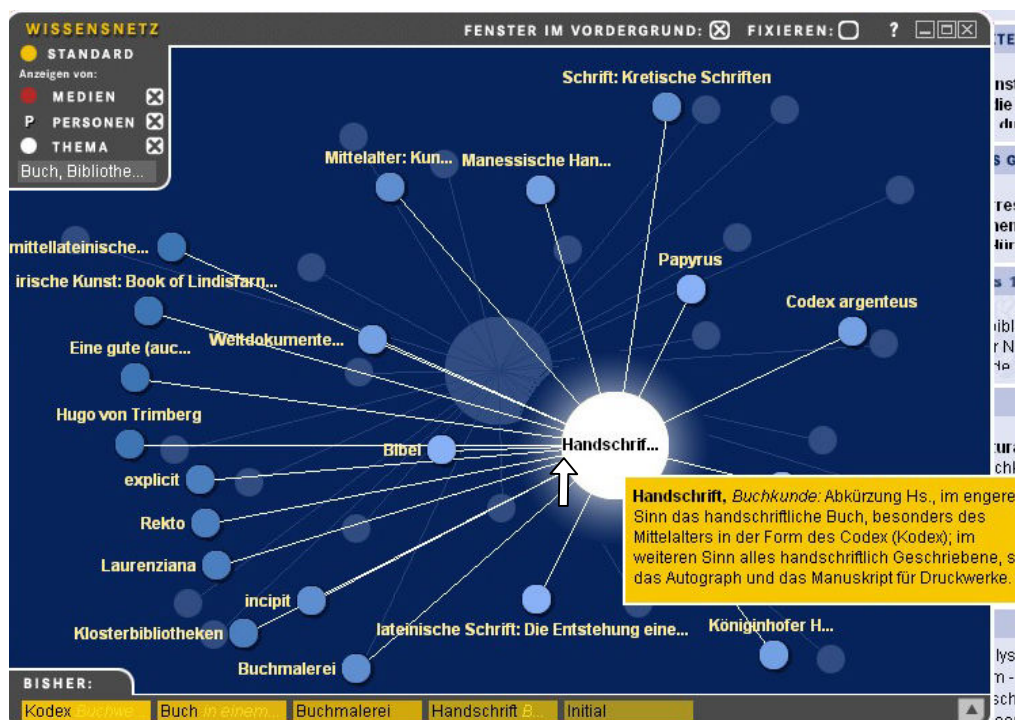


Abb. 11 Brockhaus – Vorschau auf ein neues Begriffsnetz

Da alle Einträge im Netz Lemmata sind und diese auch aus Phrasen beziehungsweise längeren Artikelüberschriften bestehen können, kann es aus Platzmangel im Anzeigefenster dazu kommen, dass nicht der ganze Eintrag angezeigt wird. Bei längeren Einträgen wird kein Umbruch vorgenommen. Zum Beispiel wird der Eintrag „Buchdruck: Eine Technik v...“ nicht vollständig dargestellt:

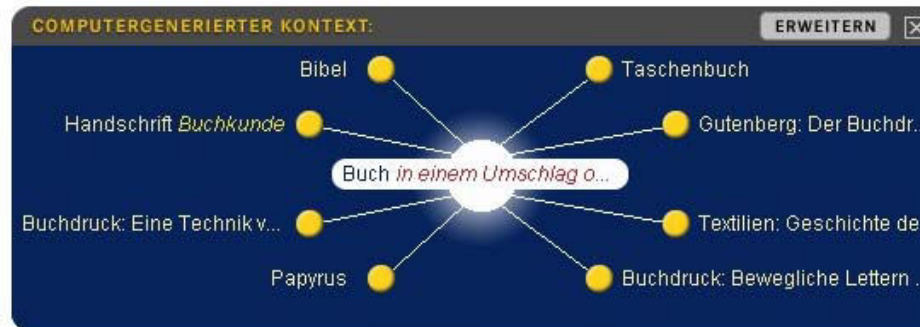


Abb. 12 Brockhaus – Wissensnetz mit nicht vollständig angezeigten Einträgen

Durch Anfahren des Eintrags „Buchdruck: Eine Technik v...“ mit der Maus wird er vervollständigt zu „Buchdruck: Eine Technik verändert die Welt“:

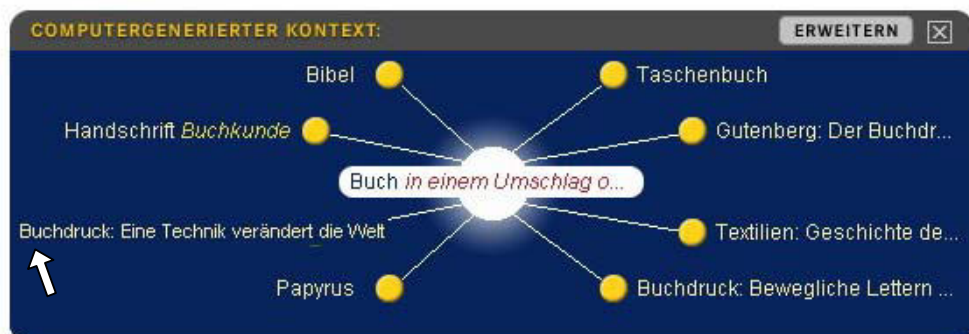


Abb. 13 Brockhaus – Wissensnetz mit vollständig angezeigtem Eintrag

Durch die strikte Begrenzung der im Netz angezeigten Terme auf 8 beziehungsweise 24, ist das Netz immer sehr übersichtlich und es kommt nicht zu Überlappungen.

Bei der Suche mit einem Synonym erfolgt die Verweisung auf die Vorzugsbenennung (also auf das Lemma) nicht automatisch, sondern nach Auswahl durch den Nutzer. Beispielsweise wird durch die Suche nach „Inkunabeln“ der entsprechende Artikel aufgerufen, der auf die Vorzugsbenennung „Wiegendrucke“ innerhalb des Artikels über einen Link hinweist.

Durch Verfolgung des Links kann der Nutzer zu diesem Artikel gelangen. Alternativ kann derselbe Artikel auch über die Trefferliste ausgewählt werden:

The screenshot displays the 'DER BROCKHAUS multimedial 2005' interface. On the left, a search bar contains 'Inkunabeln' with a 'GO!' button. Below it, a 'GESAMT' (Total) section shows a list of search results, including '100% Inkunabeln', '98% Wiegendrucke', and '38% Buchdruck: Eine Technik ver...'. A red arrow points to the '100% Inkunabeln' entry. The main content area features the title 'Inkunabeln' and a definition: '[lateinisch], Wiegendrucke.' A red arrow points to the 'Wiegendrucke' link. Below the definition is a 'COMPUTERGENERIERTER KONTEXT' (Computer-generated context) diagram showing 'Inkunabeln' at the center, connected to related terms like 'Gutenberg Johannes', 'Holzschnitt', 'Gutenberg: Der Buchdr...', 'Zimelie', 'Gutenberg: Erfindung des Buch...', 'Buchdruck: Bewegliche Lettern v...', 'Frühdruck', and 'Buchdruck: Eine Technik v...'. The right sidebar includes a 'KONTEXT' section with 'WEITERE ARTIKEL' (Further articles) and 'Brockhaus 1908' (Brockhaus 1908) section. The bottom left shows a 'WISSENSNETZ' (Knowledge network) section with 'THEMEN' (Topics) and 'Statistik' (Statistics).

Abb. 14 Brockhaus – Anzeige des Artikels zum Synonym „Inkunabeln“

Durch Auswahl der verlinkten Verweisung zur Vorzugsbenennung wird der Artikel aufgerufen. Dieser Artikel enthält möglicherweise zusätzlich zum computergenerierten Wissensnetz weitere ebenfalls verlinkte Hinweise zu Artikeln aus dem thematischen Umfeld, wie beispielsweise hier im Artikel „Wiegendrucke“ der Hinweis auf einen Übersichtsartikel:

The screenshot displays the 'DER BROCKHAUS multimedial 2005' interface. At the top, there's a navigation bar with tabs like 'LEXIKON-SUCHE', 'PROFI-SUCHE', 'LEXIKON', 'THEMEN', 'ZEITLEISTE', 'ATLAS', 'EXTRAS', 'LERNEN', 'PERSONLICH', and 'INTERNET'. A search bar on the left contains 'Inkunabeln' and a 'GO!' button. Below the search bar, a 'GESAMT SUCHERGEBNIS' section shows a list of results sorted by 'RELEVANZ'. The first result is '100% Inkunabeln' with a list of items including 'Wiegendrucke' (98%), 'Buchdruck: Eine Technik ver...' (36%), 'Alba Iula' (20%), 'Ambrosiana' (20%), 'Bayerische Staatsbibliothek' (20%), 'Buch in einem Umschlag od...' (20%), 'Drucktechnik: Von der Schri...' (20%), 'Hereford' (20%), 'Kremsmünster' (20%), and 'Levine Sherrie' (20%). A 'Suchergebnisse im Detail' button is below the list. To the right of the search results, the article 'Wiegendrucke' is displayed. It has a title 'Wiegendrucke' and a sub-header 'Wiegendrucke'. The main text describes the history of printing from the 15th to the 16th century. A 'Hier finden Sie in Überblicksartikeln weiterführende Informationen:' section lists 'Buchdruck: Eine Technik verändert die Welt' with a red arrow pointing to it. Below this, a 'COMPUTERGENERIERTER KONTEXT:' diagram shows 'Wiegendrucke' at the center, connected to nodes like 'Gutenberg, Johannes', 'Holzschnitt', 'Gutenberg: Der Buchdr...', 'Zimelie', 'Gutenberg: Erfindung des Buch...', 'Buchdruck: Bewegliche Lettern v...', 'Frühdruck', and 'Buchdruck: Eine Technik v...'. On the far right, there's a 'POWER-WEBSUCHE' section with 'Wiegendrucke' and 'OPTIONEN'. Below that, a 'KONTEXT' section lists 'PREMIUMARTIKEL' and 'WEITERE ARTIKEL'. At the bottom, there's a 'MAPPEN' section with 'BAUHAUS' and 'BEARBEITEN' buttons.

Abb. 15 Brockhaus – Anzeige des Artikels „Wiegendrucke“

Über die Einträge im semantischen Netz ist durch die Verlinkung ein leicht und intuitiv zu bedienendes Browsing durch die Artikel möglich. Brockhaus beschreibt dies als Möglichkeit „beinahe spielerisch zu verwandten Themen zu wechseln und sich so einen ganz eigenen Weg durch den Wissensdschungel zu bahnen.“¹⁷² Durch das Angebot an Artikeln aus dem Kontext einer Anfrage kann das Interesse des Suchenden auf andere Zusammenhänge und Einträge gelenkt werden. Ergänzt wird das Kontextangebot durch verlinkte Hinweise und verlinkte Begriffe in den Artikeln selbst. Im Folgenden wird ein Beispiel für eine Navigationsabfolge mit Hilfe des visualisierten und durch Interaktion zu bedienenden semantischen Netzes vorgestellt. Ausgehend vom Artikel „Wiegendrucke“ könnte beispielsweise ein Interesse am weiteren thematischen Umfeld des Buchdrucks bestehen. Inspiriert durch die Angaben im Netz kann per Mausklick der Artikel „Gutenberg: Der Buchdruck und die Folgen“ ausgewählt werden:

¹⁷² „Hilfetext im Brockhaus multimedial 2005 premium – Wissensnetz“

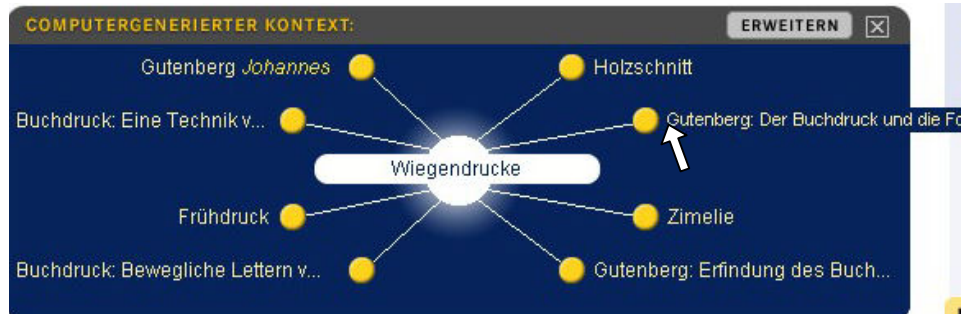


Abb. 16 Brockhaus – Auswahl des Eintrags „Gutenberg: Der Buchdruck und die Folgen“

Die Auswahl führt zu folgendem Artikel:

Abb. 17 Brockhaus – Auswahl des Eintrags „Schrift“

Das neue Netz, das zum ausgewählten Artikel aufgerufen wird, enthält das Lemma entsprechend im Zentrum der grafischen Darstellung. Das Lemma, von dem aus auf diesen Artikel zugegriffen wurde, ist zur Unterscheidung der anderen Netzeinträge in einer helleren Schriftfarbe angegeben und durch eine gestrichelte Linie mit dem Begriff des aktuell aufgerufenen Artikels verbunden. Durch den Kontext des neuen Netzes, und vielleicht

auch angeregt durch die Rezeption des Artikels, wird möglicherweise das Interesse an „Schrift“ geweckt und der Artikel „Schrift, ein System grafischer Zeichen“ aufgerufen:

The screenshot displays the Brockhaus online encyclopedia interface. The top navigation bar includes 'LEXIKON-SUCHE', 'PROFI-SUCHE', and various menu options like 'LEXIKON', 'THEMEN', 'ZEITLEISTE', 'ATLAS', 'EXTRAS', 'LERNEN', 'PERSONLICH', and 'INTERNET'. The search bar on the left shows 'Inkunabeln' with a 'GO!' button. Below the search bar, a list of search results is visible, including '100% Inkunabeln', '98% Wiegendrucke', and '38% Buchdruck: Eine Technik ver...'. The main content area is titled 'Schrift' and contains a definition: 'Schrift [althochdeutsch scrift, unter dem Einfluss von lateinisch scriptum zu althochdeutsch scriban, vergleiche Schreiben], ein System grafischer Zeichen, die zum Zweck menschlicher Kommunikation verwendet und durch Zeichnen, Malen, Einkerben, Ritzten oder Ähnliches auf feste und dauerhafte Beschreibstoffe (Stein, Rinde, Leder, Holz-, Ton- und Wachstafeln, Papyrus, Pergament, Papier u. a.) hervorgebracht werden. Im modernen Sprachgebrauch ist Schrift eine allgemeine Bezeichnung für eine Form oder ein Verfahren der Aufzeichnung oder Einprägung von Information (digitaler und analoger) auf oder in einen Träger, häufig in Zusammensetzungen verwendet; z. B. Klerschrift, Magnetschrift, Zackschrift (beim Lichttonverfahren), Wechseltaktschrift für binäre Aufzeichnung bei Magnetschichtspeichern.' Below the definition, a 'COMPUTERGENERIERTER KONTEXT' diagram is shown, featuring a central node 'Schrift ein System grafische...' connected to various related terms: 'Knochenbrüche: Formen und Be...', 'Buchdruck ein Druckverfahren (...)', 'Antiqua (Litera antiqua)...', 'Runen: Geheimnis und Geraune', 'ägyptische Schrift', 'Schrift: Festhalten von Sprache', 'Gutenberg: Der Buchdruck und die Folgen', and 'Runen'. The right sidebar contains additional information, including 'KONTEXT', 'PREMIUMARTIKEL', 'HÄTTEN SIE'S GEWUSST!', 'LITERATUR', 'MAPPEN', and 'NOTIZEN'.

Abb. 18 Brockhaus – Auswahl des Eintrags „Runen: Geheimnis und Geraune“

So kann entsprechend weiter durch das Netz navigiert werden, beispielsweise durch Aufrufen des Artikels „Runen: Geheimnis und Geraune“, aus dessen Netz wiederum ein weiterer Artikel aus dem Umfeld aufgerufen werden kann:

The screenshot displays the 'DER BROCKHAUS multimedia 2005' interface. The top navigation bar includes 'LEXIKON-SUCHE', 'PROFI-SUCHE', and various icons for navigation and search. The left sidebar shows search results for 'Wiegendrucke' and a list of related items. The main content area features the article 'Runen: Geheimnis und Geraune' with a detailed text block and a 'COMPUTERGENERIERTER KONTEXT' diagram. The diagram illustrates the relationship between 'Runen' and other concepts like 'Schrift: Kretische Schriften', 'Testament Recht', 'Morsealphabet', 'z, Z Sprache', 'Schrift ein System grafischer Zeichen', 'Normannen', and 'Wikingen Bezeichnung für die Bewohner Nordeuropas im 9.-11. Jh.'. The right sidebar contains sections for 'KONTEXT', 'WEITERE ARTIKEL', 'LITERATUR', 'MAPPEN', and 'NOTIZEN'.

Abb. 19 Brockhaus – Auswahl des Eintrags „Wikingen“

Auf diese Weise kann entlang der Einträge im semantischen Netz durch die Artikel navigiert werden. Durch Rezeption der Artikel und inspiriert durch das Angebot des Begriffskontextes kann sich das Suchinteresse im Laufe der Suche verändern und beispielsweise von „Wiegendrucke“ hin zu „Wikinger“ führen:



Abb. 20 Brockhaus – Anzeige des Artikels „Wikinger“

Dies ist eine optimale Unterstützung des Suchprozesses, wie es in Kapitel 2.4.1 beschrieben wurde. Die Angaben in den Artikeln können direkt rezipiert und weitere Informationen aus dem Umfeld durch das semantische Netz aufgerufen werden. Auf diese Weise kann ein Suchkreislauf in Gang gehalten werden, der zu immer weiteren Fragen und unmittelbar erfolgreicher Beantwortung führt.

5.2 WordSurfer

Die Firma *BiblioMondo* ist Hersteller von Bibliotheksinformationssystemen¹⁷³ und bietet seit 2003¹⁷⁴ über ihr System *ZONES* die Retrievalfunktion *WordSurfer* an¹⁷⁵. Innerhalb der sogenannten *SuchZone*, der Suchoberfläche

¹⁷³ „Homepage der Firma BiblioMondo“

¹⁷⁴ Vgl. Meldung aus dem Jahr 2003: BiblioMondo: ZONE-1 als neues Produkt. In: ABI-Technik 23 (2003) 1, S. 65

¹⁷⁵ Vgl. „Broschüre zum System ZONES der Firma BiblioMondo“, S. 4

für den OPAC¹⁷⁶ der Bibliothek, wird über die Suchfunktion *WordSurfer* die Möglichkeit geboten, sich semantische Begriffszusammenhänge grafisch visualisiert anzeigen zu lassen und Wörter aus dem Netz für die Suche zu übernehmen. Das Netz wird durch Termclustering generiert und um Begriffe aus Wörterbüchern ergänzt.

Für einen Suchvorgang im OPAC der Bibliothek stehen mit *SuchZone* insgesamt vier Suchoberflächen zur Verfügung:¹⁷⁷ „Indexsuche“, „Schnellsuche“ und „Erweiterte Suche“ mit den üblichen Suchfunktionen, wie boolesches Verknüpfen von Suchbegriffen, feldspezifisches Suchen, Trunkierung, Maskierung. Mit der vierten Suchoberfläche, *WordSurfer*, wird das semantische Umfeld von Suchbegriffen visualisiert. *WordSurfer* ist von den anderen Suchoberflächen unabhängig, das heißt eine Frage, die zum Beispiel über die Indexsuche gestartet wurde, kann nicht in *WordSurfer* fortgesetzt, sondern muss dort neu gestellt werden. Begriffe können nicht mit booleschen Operatoren verknüpft oder kategorienspezifisch gesucht werden.

Das Begriffsnetz in *WordSurfer* wird durch Termclustering, Verwendung eines Thesaurus, Verwendung von Wörterbüchern zur Übersetzung und durch eine Fuzzy-Suche generiert. Für das Clustering werden die bibliographischen Daten aus dem Index statistisch indexiert und über eine Gewichtung Term-Ähnlichkeiten bestimmt. Zu dem Suchbegriff einer Anfrage werden dann die über das Termclustering als ähnlich ermittelten Begriffe ausgegeben. Die Gewichtungswerte der statistischen Indexierung werden zusätzlich für ein Relevance Ranking der Ergebnismenge genutzt. Die aus dem Thesaurus ermittelten Wörter sind Synonyme zum Suchbegriff. Dabei ist zu beachten, dass es sich nicht um einen dokumentationssprachlichen Thesaurus handelt, sondern um einen Thesaurus im Sinne eines Synonymwörterbuchs, das „zu einem Wort bedeutungsverwandte Begriffsfelder“¹⁷⁸ angibt. Einfache Begriffe werden durch hinterlegte Wörterbücher ins Englische, Französische, Holländische und Spanische übersetzt. Die Fuzzy-Termgenerierung erfolgt in Echtzeit durch einen Abgleich des Suchwortes mit den Indexeinträgen der Datenbank. Darüber werden Schreibvarianten für das Netz ermittelt.¹⁷⁹

Die Suchoberfläche von *WordSurfer* ist in zwei Hälften unterteilt. Links befinden sich die Eingabezeile, die Navigationsleiste und das Fenster für die

¹⁷⁶ Abkürzung für Online Public Access Catalogue

¹⁷⁷ Diese und folgende Angaben zu *SuchZone* und *WordSurfer* beruhen auf Erkundungen durch die Verfasserin, soweit dies nicht anders gekennzeichnet ist. Die Angaben und Screenshots stützen sich auf die ZONES-Anwendung in der Stadtbibliothek Köln. Siehe dazu „ZONES-Web-OPAC der Stadtbibliothek Köln“, Zugang über https://katalog.stbib-koeln.de/alswww2.dll/APS_OPAC&Style=Portal2. Der Online-Zugang zum OPAC befindet sich noch in der Testphase.

Weitere Anwenderbibliotheken von BiblioMondo-Software in Deutschland sind auf der Homepage der Firma angegeben. Siehe dazu „Homepage der Firma BiblioMondo – Kunden in Deutschland“. Zu keiner weiteren der dort genannten Bibliotheken steht ein Online-Zugang zur ZONES-Anwendung und somit zu *WordSurfer* zur Verfügung [Stand: März 2005].

¹⁷⁸ Gödert: Sprache und Bedeutung, WS 2003/04, S. 11

¹⁷⁹ Die Angaben zur Termgenerierung für das semantische Netz in *WordSurfer* basieren auf Produktinformationen von BiblioMondo. Siehe dazu „Vertrauliche Angaben zum Produkt *WordSurfer* der Firma BiblioMondo“, S. 1f.

grafische Darstellung des Begriffsnetzes. Rechts werden die bibliographischen Angaben der gefundenen Titel nach fallender Relevanz aufgelistet. Nach Eingabe eines Suchwortes in die Eingabezeile und Drücken des Häkchen-Buttons wird das Netz aufgebaut. Die ermittelten Begriffe aus dem semantischen Umfeld werden um das Suchwort im Zentrum in zufälliger Position angeordnet. Parallel dazu werden in der rechten Bildschirmhälfte die gefundenen Titeldaten aufgelistet. Die Generierungsart der Begriffe im Netz wird durch unterschiedliche Kolorierung visualisiert: die Begriffe aus dem Termclustering, als „Angrenzende Wörter“ bezeichnet, in der Farbe Grün, die Begriffe aus dem Thesaurus („Verwandte Begriffe“) in der Farbe Rot, die Begriffe aus der Fuzzy-Suche („Alternative Schreibweise“) in der Farbe Gelb und die Begriffe aus den Übersetzungswörterbüchern („Übersetzte Begriffe“) in der Farbe Hellblau. Beispielsweise führt eine Suche nach „Buch“ zu folgenden Angaben:

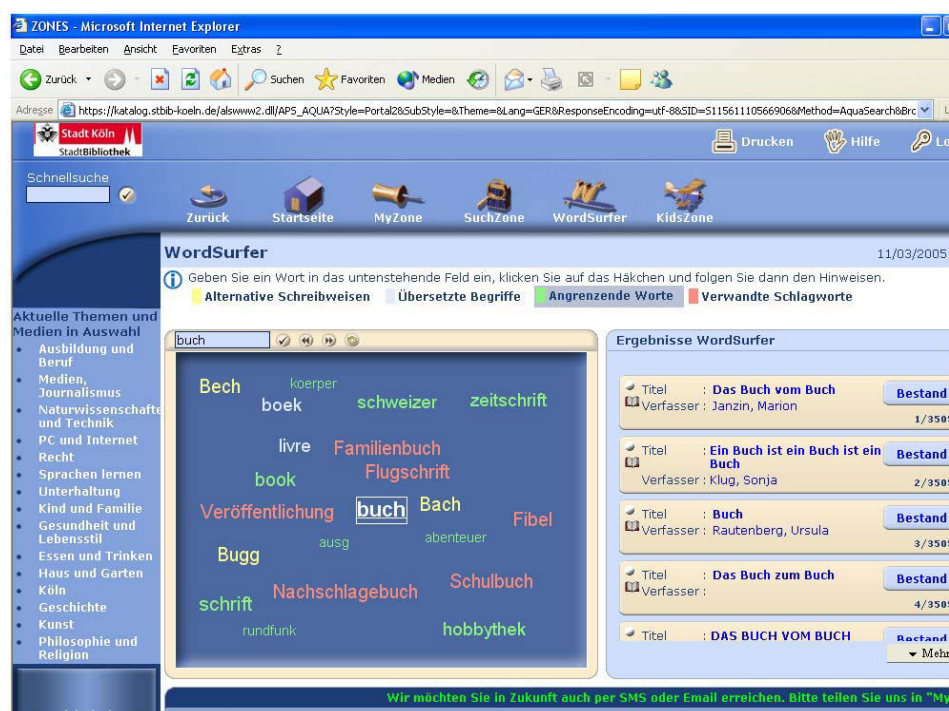


Abb. 21 N.N. 2005a – Suche nach „Buch“

Die verschiedenen Relevanzgrade der Terme werden durch eine unterschiedliche Schriftgröße kenntlich gemacht. Dabei bedeutet eine größere Schrift eine höhere Relevanz, also eine höhere Gewichtung. Dies soll dem Nutzer einen visuellen Eindruck von der Wichtigkeit der generierten Terme geben. Unbedeutend ist es, wie nahe oder weit entfernt Begriffe aus dem Netz zum Suchterm stehen. Ebenso irrelevant ist die Groß- und Kleinschreibung der Wörter.

Durch Mouse-Over über einen Netzeintrag wird das entsprechende Wort eingerahmt, die Schriftfarbe ändert sich, eine Verbindungslinie zum Begriff im Zentrum wird gezogen und in der Legende zur Termgenerierung wird markiert, um welche Termart es sich handelt. Diese kombinierte Kenntlichmachung des ausgewählten Begriffs aus dem Netz verhindert ein versehentliches Anklicken eines nicht gewünschten Terms. Im folgenden Beispiel ist der Term „Veröffentlichung“ aus dem semantischen Netz zu „Buch“ mit der Maus angefahren:

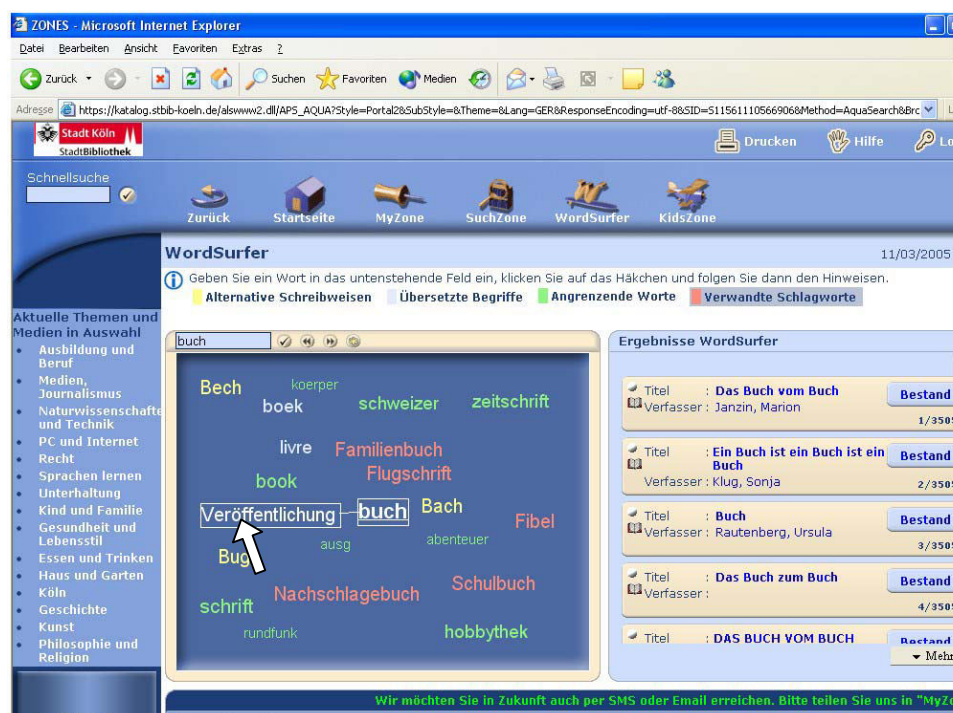


Abb. 22 N.N. 2005b – Auswahl des Eintrags „Veröffentlichung“ im Netz zu „Buch“

Alle Benennungen im Netz sind Hotwords und können durch Anklicken mit der Maus in die Suche übernommen werden. Wird ein Begriff aus dem Netz angeklickt, baut sich das Netz und parallel dazu die Ergebnismenge der bibliographischen Titeldaten neu auf. Der Übergang vom alten zum neuen Netz ist animiert. Einige Terme werden dabei größer, andere kleiner, manche verschwinden und neue Terme kommen hinzu. Insgesamt soll der Eindruck entstehen, dass sich die Wörter aus der Tiefe des Bildschirms in den Vordergrund schieben. Durch die unterschiedlichen Schriftgrößen soll die Illusion einer Perspektive entstehen. *BiblioMondo* beschreibt die Navigation durch das Netz folgendermaßen: „If you click on a ‚found‘ word, you ‚move‘ through the ‚virtual‘ space to this word, such that some words re-

cede to the distance (grow smaller or even disappear) while some words appear to move closer, increasing in size.“¹⁸⁰

Die aus dem semantischen Netz aufeinanderfolgend per Mausklick ausgewählten Terme werden zu einer Suchanfrage kombiniert. Dabei entwickelt sich eine Suche ausgehend vom eingetippten Suchwort. Zu diesem Term wird das semantische Netz aufgebaut, aus dem ein weiterer Term zur Suchanfrage ausgewählt werden kann. Zu diesem Term wird erneut ein semantisches Netz aufgebaut, aus dem wiederum ein weiterer Begriff gewählt werden kann und die Suchanfrage ergänzt. So kann die Suchanfrage mit jedem Navigationsschritt weiterentwickelt werden. Mit jedem neu ausgewählten Eintrag ändert sich nicht nur jeweils das Netz, sondern entsprechend zur Suchanfrage auch die Ergebnismenge. Dabei werden die Suchbegriffe nicht zu einer booleschen Anfrage zusammengesetzt, sondern die Ergebnismenge wird durch die unterschiedlichen Gewichtungswerte der ausgewählten Terme ermittelt. Eine Suchabfolge durch Navigation im semantischen Netz kann ausgehend von einer Suche nach „Buchdruck“ beispielsweise wie folgt ablaufen:

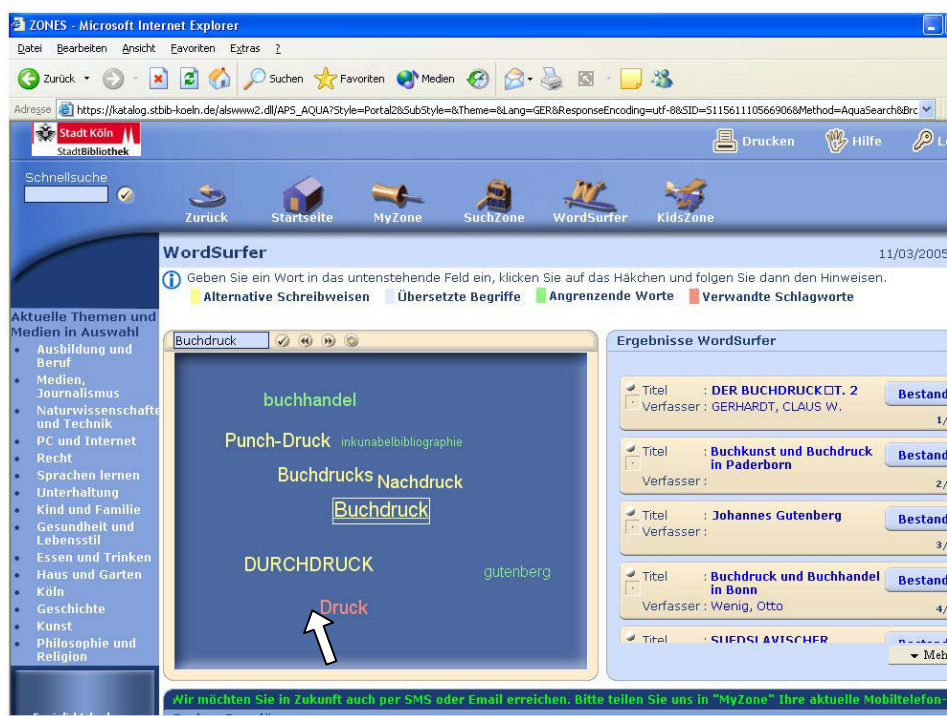


Abb. 23 N.N. 2005c – Suche nach „Buchdruck“, Auswahl des Eintrags „Druck“

¹⁸⁰ „Vertrauliche Angaben zum Produkt WordSurfer der Firma BiblioMondo“, S. 2

Die Suche kann durch Termauswahl im Netz um „Druck“ erweitert werden:

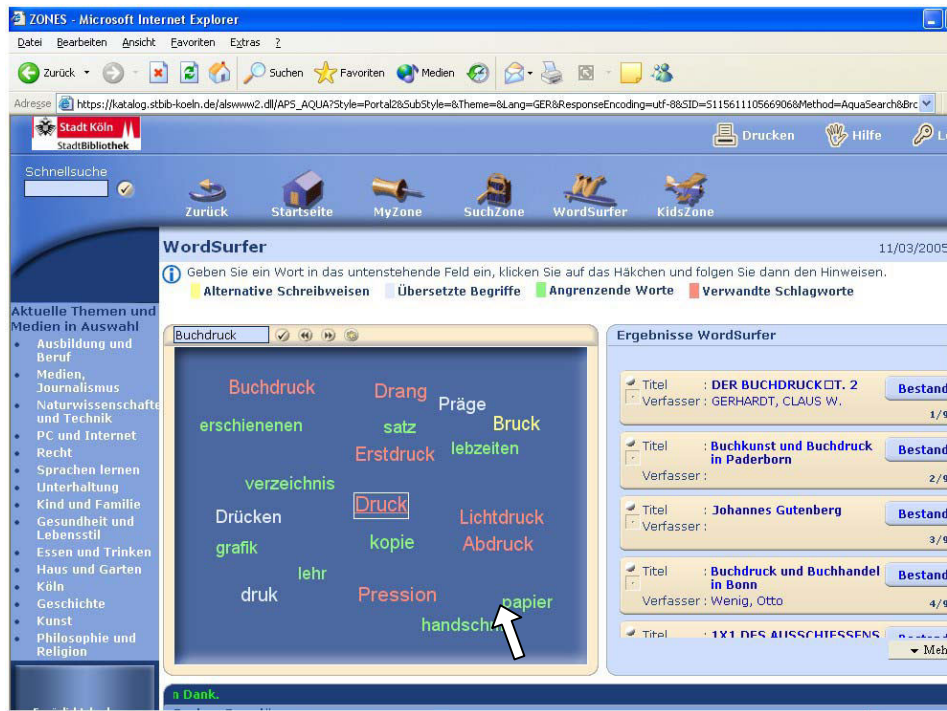


Abb. 24 N.N. 2005d – Auswahl des Eintrags „Papier“ im Netz zu „Druck“

Im Netz wird der im Zusammenhang mit „Druck“ stehende Term „Papier“ ausgegeben und kann in einem weiteren Navigationsschritt ausgewählt werden. Die Suche wird um diesen weiteren Begriff ergänzt:

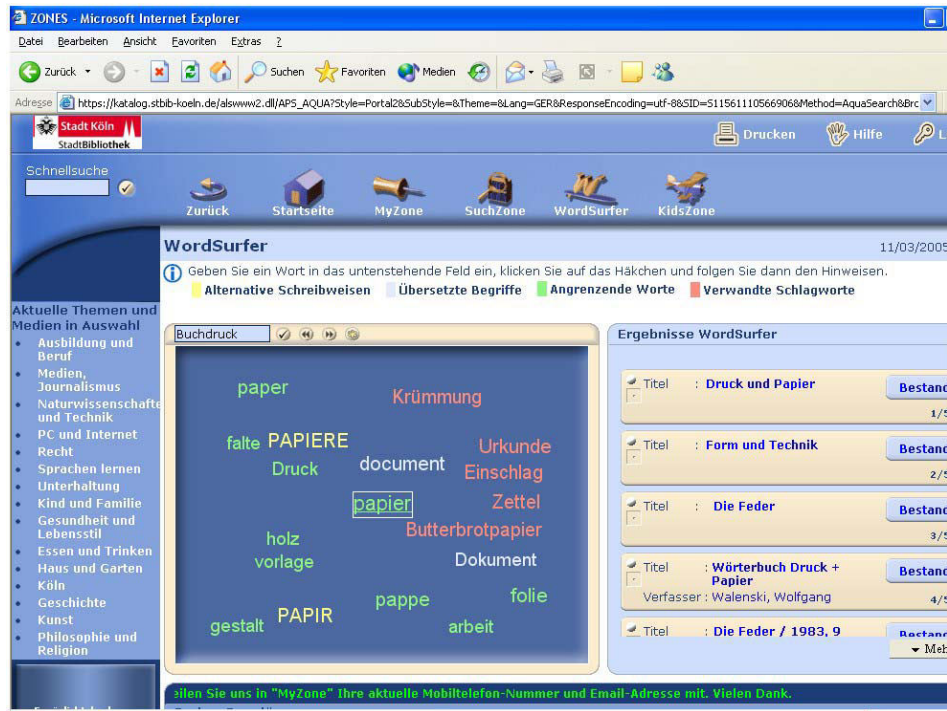


Abb. 25 N.N. 2005e – Netz zu „Papier“

Für die Generierung des Begriffsnetzes macht es keinen Unterschied, ob der Term aus dem ersten Schritt durch das manuell eingegebene Suchwort generiert, oder aus einem x-ten Navigationsschritt aufgerufen wird – es werden immer die gleichen Begriffe zum Term im Zentrum erzeugt. Über die intuitiv bedienbare Suchoberfläche können die Terme leicht ausgewählt und zu einer Suche zusammengestellt werden. Die Trefferbildung über die Gewichtungswerte bietet eine gute Alternative zur booleschen Suchmöglichkeit der anderen Suchoberflächen in der *SuchZone*.

Eine neue Suche kann durch den Sternchen-Button in der Navigationsleiste begonnen werden. Das aktuell aufgerufene Netz und die Treffermenge werden zurückgesetzt und es kann ein neues Suchwort in die Eingabezeile eingegeben werden. Mit den Pfeiltasten kann zwischen einzelnen Suchschritten zurück und vor navigiert werden. Hier die Abbildung der Navigationsleiste:



Abb. 26 N.N. 2005f – Navigationsleiste

Die Darstellung des Netzes zu einfachen Begriffen ist sehr übersichtlich. Dagegen kann es bei Komposita und Begriffen zu denen die meisten Netzeinträge durch die Fuzzy-Termgenerierung erzeugt werden, in dem kleinen Anzeigefenster zu Überlappungen kommen. Die Termauswahl wird dadurch erschwert, denn es gibt keine Funktion, die die Begriffe lesbar macht, wie beispielsweise Begriffe durch Mouse-Over in den Vordergrund zu rücken. Ein Beispiel für eine derartige Überlappung geschieht bei der Suche nach „Bibliothekswesen“ und der anschließenden Auswahl des Begriffs „Bibliothekswelt“ aus dem Netz:

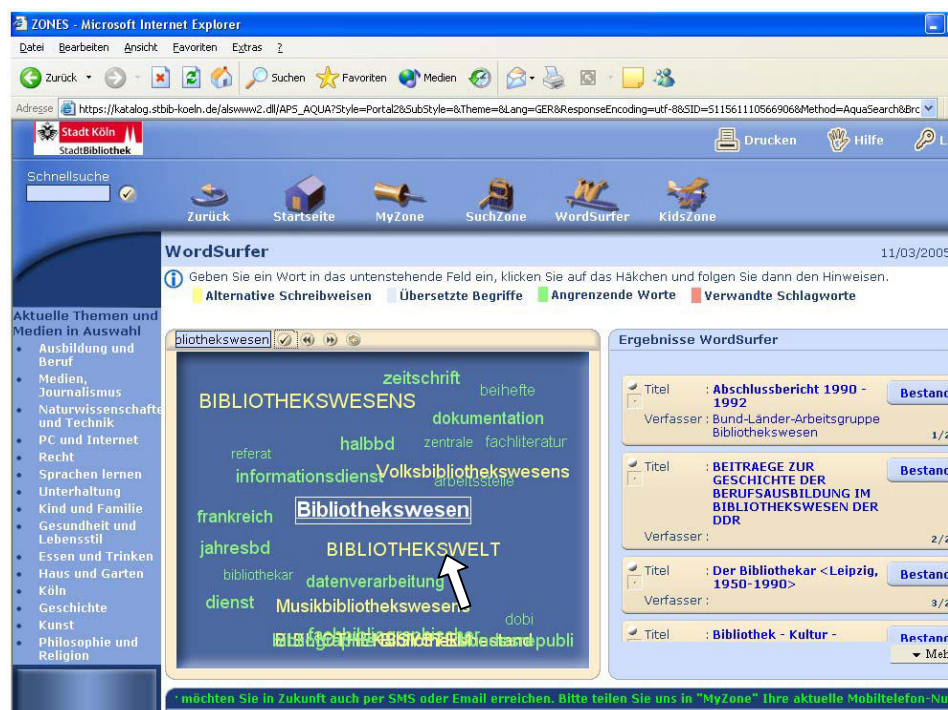


Abb. 27 N.N. 2005g – Suche nach „Bibliothekswesen“, Auswahl des Eintrags „Bibliothekswelt“

Anzeige des Netzes zu „Bibliothekswelt“ mit starken Termüberlappungen:

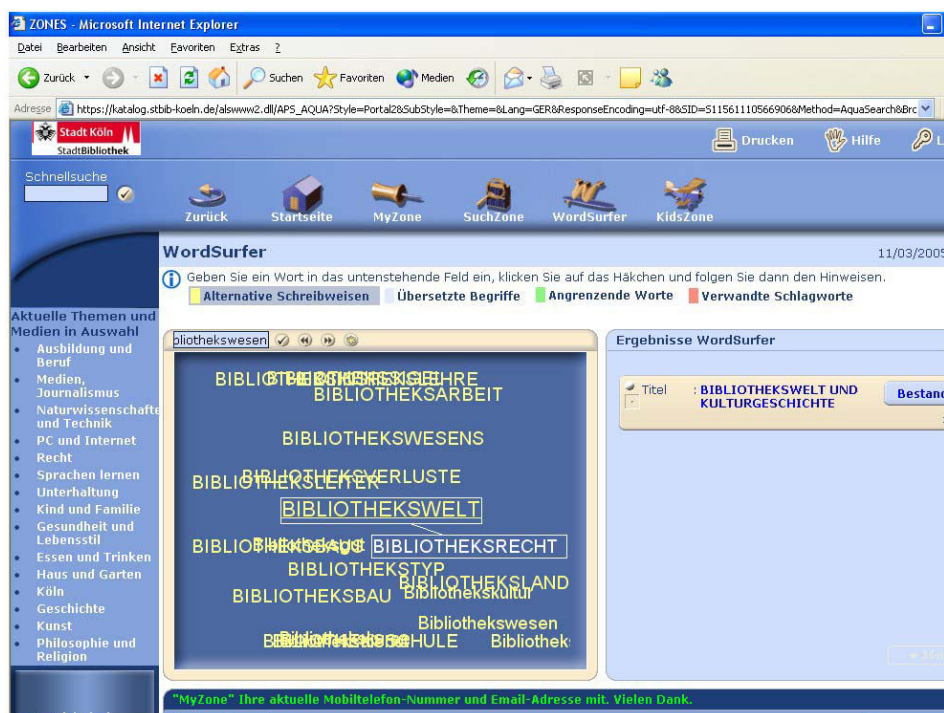


Abb. 28 N.N. 2005h – Anzeige des Terms „Bibliothekswelt“ ausgewählt aus dem Netz zu „Bibliothekswesen“ (Überlappung)

Neben der Möglichkeit, die Suche mit einem einzelnen Suchbegriff zu starten, können auch mehrere Begriffe kombiniert gesucht werden. So erzeugt beispielsweise die Eingabe von „Katalog Bibliothek“ ein anderes Netz und eine andere Treffermenge als die Suche nach den einzelnen Begriffen „Katalog“ oder „Bibliothek“. Durch die Suche von „Katalog Bibliothek“ wird erreicht, dass in der Ergebnismenge Titel zu Katalogen aus dem bibliothekarischen Bereich gefunden werden und nicht beispielsweise Titel zu Ausstellungskatalogen aus dem Themenbereich Kunst. Eine Auswahl über einen Dialog, welche Katalogart gesucht wird, wird nicht angeboten und eine Disambiguierung durch Homonymenzusätze bei Schlagwörtern ist nicht möglich. Beispielsweise führt eine Suche nach dem Homonym „Krebs“ zu Titeln mit den Schlagwortangaben „Krebs <Medizin>“ und „Krebs <Sternbild>“, deren Zusatz jedoch nicht im Netz ausgegeben werden, da keine Phrasen indexiert werden:

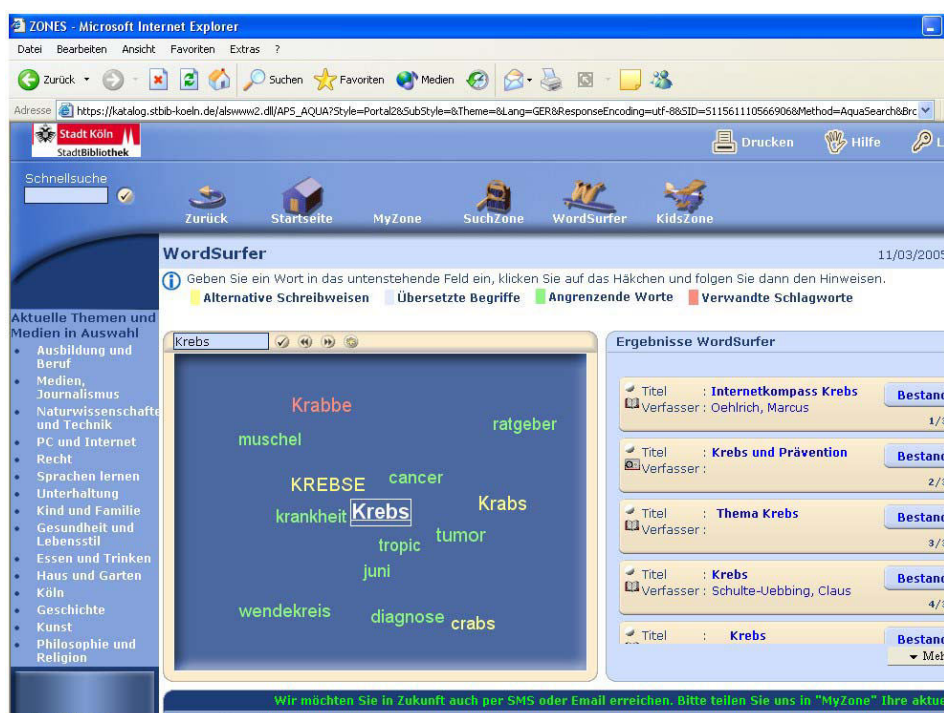


Abb. 29 N.N. 2005i – Suche nach dem Homonym „Krebs“

Durch das Clustering, die Fuzzy-Generierung und das Synonymwörterbuch werden lediglich weitere Terme zur Verfügung gestellt, so dass durch Ergänzung von Begriffen aus dem Kontext die Suche auf ein bestimmtes Themengebiet gelenkt werden kann.

Eine Suche nach der Kombination „Krebs Sternbild“ erzeugt eine Ergebnismenge mit Titeln an vorderster Stelle, die beide Suchterme enthalten, wie

beispielsweise Titel aus dem Bereich Astrologie mit dem Schlagwort „Krebs <Sternbild>“, gefolgt von Titeln, die einem der Suchkriterien entsprechen, wie Werke zu Krebs aus dem medizinischen Bereich und Werke zu Sternbildern aus dem Gebiet Astronomie. Die Möglichkeit, nach mehreren Begriffen kombiniert suchen zu können, erlaubt auch die Suche nach Phrasen wie „Öffentliche Bibliothek“, auch wenn nicht eine explizite Phrasensuche durchgeführt wird. Titelangaben, die die Suchterme in der gleichen Reihenfolge und mit gleichem Zeichenabstand wie in der Suchanfrage enthalten, werden höher gewichtet und in der Treffermenge demzufolge zuerst ausgegeben. Dahinter folgen Treffer, die die Suchterme in flektierter Form enthalten und danach werden Titel aufgelistet, die nur einen der beiden Terme der Anfrage aufweisen. Im Folgenden ein Ausschnitt des Netzes und der ersten 4 Treffer der Ergebnismenge:

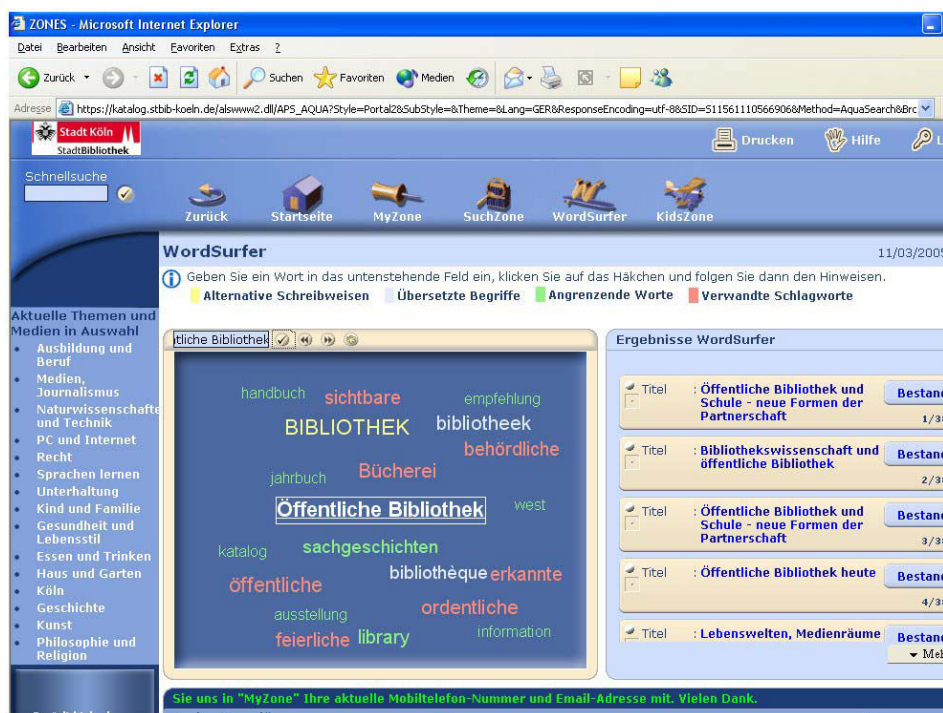


Abb. 30 N.N. 2005j – Suche nach „Öffentliche Bibliothek“

Für das Ranking der Ergebnismenge ist es problematisch, dass zum einen nicht zwischen belletristischer Literatur und Sachliteratur unterschieden wird und zum anderen, dass über die bibliographischen Angaben sowohl auf formale als auch auf inhaltliche Aspekte gemischt zugegriffen wird. Alle Angaben einer Titelaufnahme werden für die Netzgenerierung und für das Ranking der Ergebnismenge herangezogen, so dass es dadurch zu Treffern kommen kann, die vom Nutzer nicht gewünscht sind. So wird beispielsweise

und zu einer Suche nach „Tanne“ werden keine Titelangaben mit dem Wort „Nadelbaum“ gefunden.

Für Bibliotheken aus dem englischsprachigen Raum wird zusätzlich zur statistischen eine linguistische Indexierung durchgeführt. Über Anwendung eines Stemmingalgorithmus wird eine Suchanfrage automatisch auf Indexate mit gleichem Wortstamm ausgeweitet.¹⁸¹ Zusätzlich werden in der englischen Version von *WordSurfer* auch Komposita zerlegt. Das bedeutet, dass zu einer Suche nach „Rosegarden“ auch Titel gefunden werden, die „Rose“ und „Garden“ enthalten beziehungsweise eine Anfrage nach „Rose“ auch Titel mit „Rosegarden“ ermittelt.¹⁸² Teile des Kompositums werden jeweils geringer gewichtet als der eigentliche Suchterm. Für die deutsche Sprache steht dieses Angebot nicht zur Verfügung.

Verschiedene Schreibvarianten eines Suchbegriffs, die über die Fuzzy-Suche generiert werden, erweitern die Anfrage nicht automatisch. Das heißt die Ergebnismenge wird nicht um Titel ergänzt, die andere Schreibvarianten des Suchterms enthalten. Je nachdem mit welcher Variante die Suche begonnen wird, werden unterschiedliche Treffermengen erzeugt. Beispielsweise führt eine Suche nach „Tschaikowsky“ zu 49 Treffern und eine Suche nach „Tschaikowski“ nur zu 8 Titeln. Durch Auswahl alternativer Schreibweisen aus dem Netz können sie mit in die Suche einbezogen werden. Das gleiche gilt für Rechtschreibfehler, die, anders als es in der Broschüre beschrieben ist¹⁸³, ebenfalls nicht automatisch durch WordSurfers Fuzzy-Logic behoben werden, sondern auch in diesem Fall nur durch Auswahl der angebotenen Schreibvarianten aus dem Netz in die Suche einbezogen werden können. Es besteht dabei keine Möglichkeit, nach der Begriffsauswahl eines fuzzy-generierten Terms die Suche neu zu starten. Das bedeutet, dass der ausgewählte Term nur in Kombination mit dem ursprünglichen, falsch geschriebenen Term gesucht werden kann. Führt der Schreibfehler dazu, dass keine Titel gefunden werden, ist dies kein Problem. Beispielsweise führt eine Suche nach „Bucg“, gemeint ist „Buch“, zu keinem Treffer, so dass durch Auswahl des Terms „Buch“ aus dem Netz der Fehler korrigiert wird. Führt der Tippfehler jedoch zu einer Suchanfrage nach einem Begriff, der zwar nicht gemeint ist, zu dem aber Titel ermittelt werden können, führt eine Begriffsauswahl aus dem Netz nicht zu einer Suche mit dem gewünschten Term, sondern zu einer kombinierten Anfrage. Beispielsweise erzeugt die Suche nach „Bach“, gemeint ist „Buch“, durch Auswahl des Terms „Buch“ aus dem Netz eine Treffermenge, die Titel aus der Begriffskombination an vorderer Stelle auflistet, wie zum Beispiel „Das neue Bach-Blüten-Buch“:

¹⁸¹ Als Beispiel wird in den vertraulichen Angaben zum Produkt WordSurfer der Firma BiblioMondo angegeben, dass über das Suchwort „Cycle“ auch „Cycles“, „Cycling“ und „Cycled“ gefunden wird. Vgl. „Vertrauliche Angaben zum Produkt WordSurfer der Firma BiblioMondo“, S. 3

¹⁸² Das Beispiel ist den vertraulichen Angaben zum Produkt WordSurfer der Firma BiblioMondo entnommen. Vgl. „Vertrauliche Angaben zum Produkt WordSurfer der Firma BiblioMondo“, S. 1 und S. 3

¹⁸³ Vgl. „Broschüre zum System ZONES der Firma BiblioMondo“, S. 4

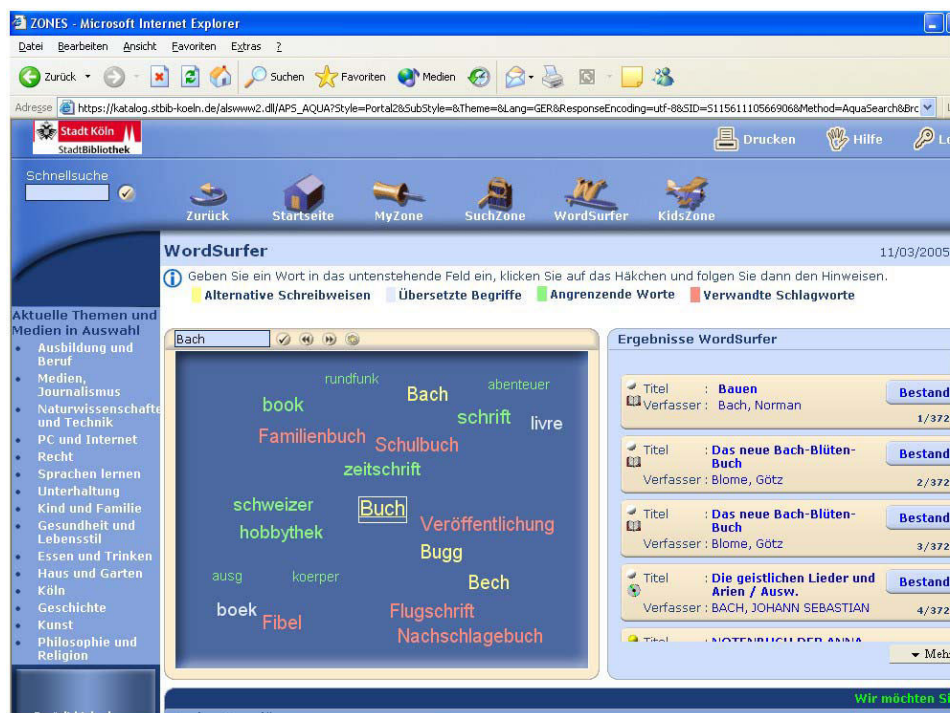


Abb. 32 N.N. 2005l – Anzeige des Terms „Buch“ ausgewählt aus dem Netz zu „Bach“

5.3 SpiderSearch

Die Firma *BOND* ist Hersteller von Bibliothekssoftware und bietet zu ihrer Software *BIBLIOTHECA2000* seit 2002/2003 den Assoziativ-OPAC *SpiderSearch* an.¹⁸⁴ „SpiderSearch ist ein Zusatzmodul zum BOND-Web-OPAC“¹⁸⁵ und bietet die Möglichkeit, Begriffe aus dem semantischen Umfeld eines Suchwortes grafisch visualisiert anzuzeigen und Wörter aus dem Begriffsnetz in eine Suche zu übernehmen. Das Netz wird vermutlich über ein Termclustering der Katalogdaten erzeugt. Angaben in der Broschüre zu *SpiderSearch*, nach deren Aussage eine Ähnlichkeitssuche aufgrund von statistisch ermittelten Gewichten erfolgt¹⁸⁶ und Erkundungen der Anwendung durch die Verfasserin legen dies nahe. Laut den Angaben auf der Homepage von *BOND* werden auch dann Titel in der Ergebnismenge ausgege-

¹⁸⁴ Es gibt sowohl Ankündigungen aus dem Jahr 2002 als auch aus dem Jahr 2003. Vgl. folgende Artikel: Spidersearch von BOND – Neue Suchvariante mit grafischem Assoziativ-OPAC. In: ABI-Technik 22 (2002) 3, S. 278 ; BOND: SpiderSearch – der neue grafische Assoziativ-OPAC. In: Bibliotheksdienst 36 (2002) 10, S. 1389f. ; Revolutionäre neue Suchvariante mit neuem grafischen Assoziativ-OPAC Spidersearch von BOND. In: B.I.T.online 5 (2002) 3, S. 264 ; BOND: Faszinierende Suchergebnisse durch visuell animierte Medienrecherche. In: ABI-Technik 23 (2003) 3, S. 247 ; BOND: Assoziativ-OPAC SpiderSearch. In: Bibliotheksdienst 37 (2003) 8/9, S. 1145 ; Assoziativ-OPAC Spidersearch von BOND. In: B.I.T.online 6 (2003) 3, S. 281

¹⁸⁵ „Broschüre zu SpiderSearch der Firma BOND“, S. 1

¹⁸⁶ Vgl. ebd., S. 2

ben, wenn Terme im Datensatz eine Schreibvariante des Suchbegriffs darstellen. Zusätzlich werden die alternativen Schreibweisen im Netz angezeigt.¹⁸⁷ Dies würde auf eine Fuzzy-Suche schließen lassen. Durch Erkundungen wurde jedoch ermittelt, dass Schreibvarianten nur dann ausgegeben werden, wenn sie aus dem Termclustering erzeugt werden; eine Fuzzy-Termgenerierung erfolgt nicht. Die aus der statistischen Indexierung der Daten ermittelten Gewichte werden für die Generierung der ähnlichen Begriffe zu einer Suchanfrage und für ein Ranking der Treffermenge nach Relevanz genutzt. Es findet keine linguistische automatische Indexierung statt.

Für eine Suche im OPAC der Bibliothek stehen drei verschiedene Suchmasken zur Verfügung.¹⁸⁸ Eine „Einfache Suche“ und eine „Profisuche“ mit folgendem Funktionsumfang: boolesche UND-Verknüpfung, Trunkierung, Maskierung, feldspezifische Suche und Registersuche nach Interessenkreisen und Systematikangaben. Die dritte, eigenständige Suchoberfläche *SpiderSearch*, ermöglicht eine grafische Anzeige des Begriffsumfeldes zu einer Suchanfrage. Eine Suche, die über die Suchmasken „Einfache Suche“ oder „Profisuche“ begonnen wird, kann nicht mit *SpiderSearch* fortgesetzt werden, sondern muss dort neu gestellt werden.

¹⁸⁷ Vgl. „Homepage der Firma BOND – zu SpiderSearch“

¹⁸⁸ Diese und folgende Angaben zur Suche in BIBLIOTHECA2000 und zur Nutzung von SpiderSearch beruhen auf Erkundungen durch die Verfasserin, soweit dies nicht anders gekennzeichnet ist. Die Angaben und Screenshots stützen sich auf die SpiderSearch-Anwendung in der Stadtbibliothek Herten. Siehe dazu „BOND-Web-OPAC der Stadtbibliothek Herten“, Zugang über <http://www.herten.de/kultur/bibliothek/index.htm> und Aufrufen des Links „Hier geht es zur Recherche im Online-Katalog der Stadtbibliothek Herten“

In Deutschland gibt es drei weitere Bibliotheken, die nach Auskunft von Bibliomondo SpiderSearch nutzen. Siehe „Anwenderbibliotheken von SpiderSearch nach Angaben der Firma BOND“ Über einen Web-OPAC können jedoch nur die Daten der Stadtbücherei Bornheim und der Stadtbibliothek Herten mit SpiderSearch angezeigt werden [Stand: März 2005]. Für die Erkundungen wurde von der Verfasserin der Katalog der Stadtbibliothek Herten ausgewählt, da er den größeren Bestand nachweist. Die SpiderSearch-Nutzung der Stadtbibliothek Reutlingen seit April 2005, vgl. dazu die Meldung „Stadtbibliothek Reutlingen: Web-Auftritt komplett mit Web-OPAC Premium von BOND realisiert. In: BuB 57 (2005) 4, S. 271, konnte für diese Arbeit nicht mehr berücksichtigt werden.

Die Suchoberfläche von *SpiderSearch* ist in zwei Hälften geteilt. Links befinden sich die Eingabezeile und das Fenster für die Anzeige der bibliographischen Angaben aus der ermittelten Treffermenge. Auf der rechten Seite wird das Begriffsnetz visualisiert. Durch Eintippen eines Suchwortes in die Eingabezeile und Anklicken des Suchen-Buttons wird das Netz aufgebaut. Die nach dem Clustering als ähnlich zur Suchanfrage ermittelten Begriffe werden um das Suchwort im Zentrum herum kreisförmig angeordnet und sind mit diesem durch strahlenförmige Linien verbunden. Parallel dazu werden in der linken Bildschirmhälfte die gefundenen Titeldaten angezeigt. Die Suche kann über Auswahl aus einem Pull-down-Menü auf den Bestand einer bestimmten Zweigstelle und auf eine Medienart eingeschränkt werden. Beispielsweise erzeugt eine Suche nach „Buch“ folgende Angaben:

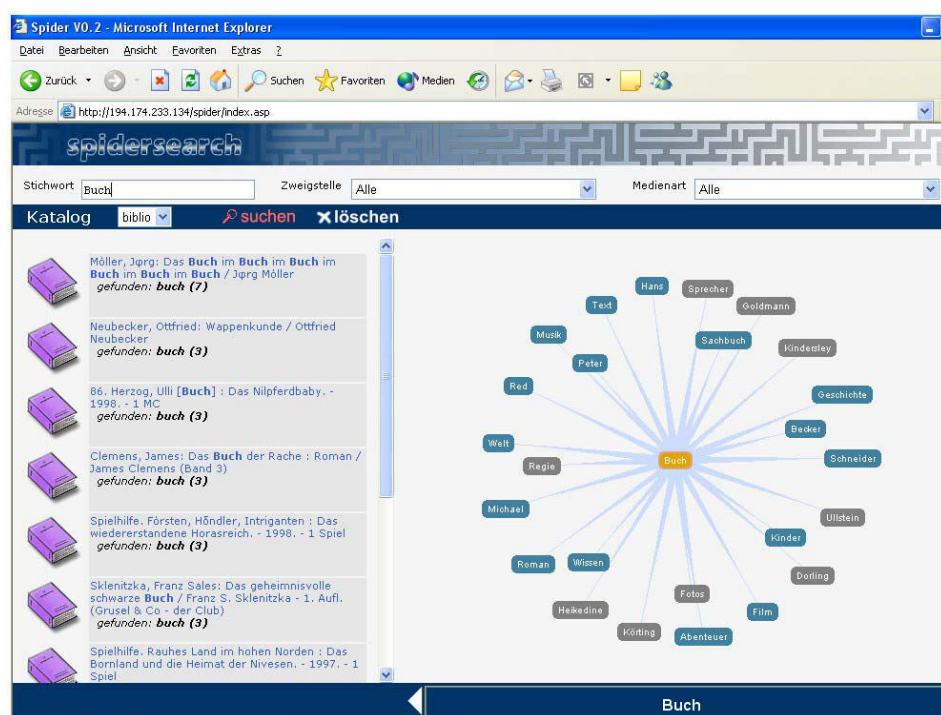


Abb. 33 N.N. 2005m – Suche nach „Buch“

In der Netzanzeige sollen laut Broschüre zu *SpiderSearch* „unterschiedliche Verwandtschaftsarten ... durch unterschiedliche Farben und Größen“¹⁸⁹ deutlich gemacht werden. Tatsächlich sind die Begriffe, abgesehen von der Wortlänge, jedoch immer gleich groß. Die Bedeutung der unterschiedlichen Farbgebung wird weder in einer Legende, noch in der Literatur oder im Hilfetext angegeben und ist auch nicht selbsterklärend.

¹⁸⁹ „Broschüre zu SpiderSearch der Firma BOND“, S. 1

Durch Mouse-Over über einen Term wird die Auswahl kenntlich gemacht, indem der Kasten, in dem der Term steht, rot eingrahmt wird. Im folgenden Beispiel ist dies der Term „Abenteuer“ im Netz zu „Buch“:

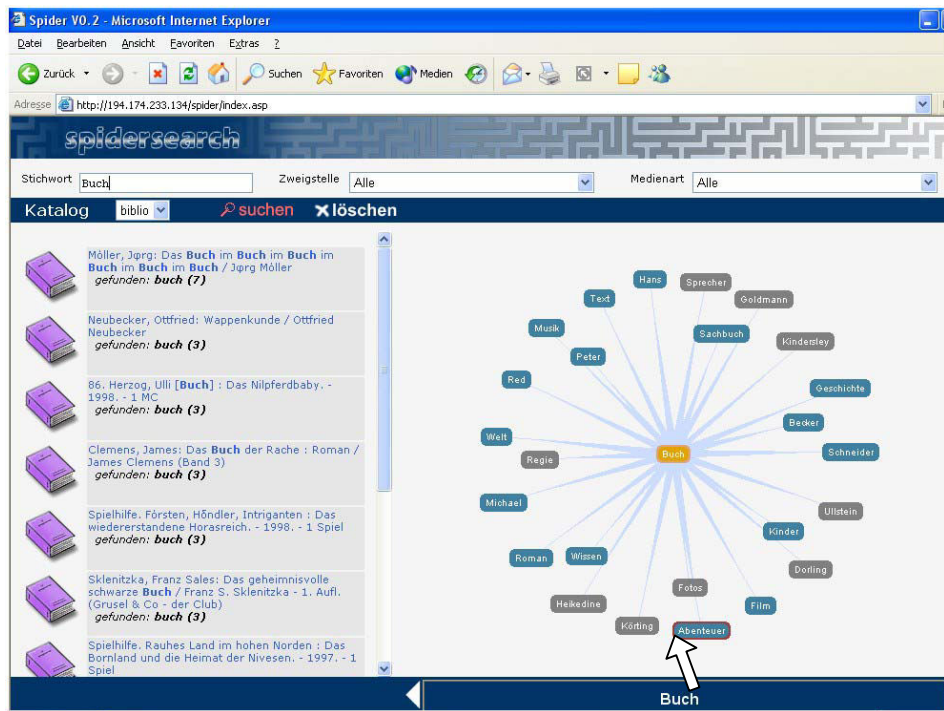


Abb. 34 N.N. 2005n – Auswahl des Eintrags „Abenteuer“ im Netz zu „Buch“

Die Darstellung der Terme im Netz ist aufgrund der begrenzten Anzahl der angezeigten Begriffsmenge sehr übersichtlich.¹⁹⁰ Sollte es dennoch einmal zu einer Begriffsüberlappung kommen, tritt der Term durch Anfahren mit der Maus in den Vordergrund und ist dadurch lesbar.

Alle Kästen im Netz, die die Begriffe enthalten, sind Hotspots, so dass sie durch Anklicken mit der Maus in eine Suche übernommen werden können. Wird ein Begriff aus dem Netz angeklickt, baut sich das Netz um diesen Term neu auf und die neue Ergebnismenge wird angezeigt. Der Aufbau des neuen Netzes wird animiert dargestellt, indem sich das neue Netz aus dem ursprünglichen Netz entwickelt. Dazu spaltet sich der ausgewählte Term ab und gruppiert neue Begriffe um sich, während das alte Netz verschwindet. Beispielsweise baut sich das Netz ausgehend vom semantischen Netz „Buch“ nach Auswahl des Begriffs „Abenteuer“ wie folgt neu auf:¹⁹¹

¹⁹⁰ Bei der Erkundung wurde nie eine so hohe Begriffsanzahl festgestellt, dass die Darstellung unübersichtlich gewesen wäre.

¹⁹¹ Der hier nur über 8 Screenshots dargestellte Vorgang verläuft in Wirklichkeit fließend.

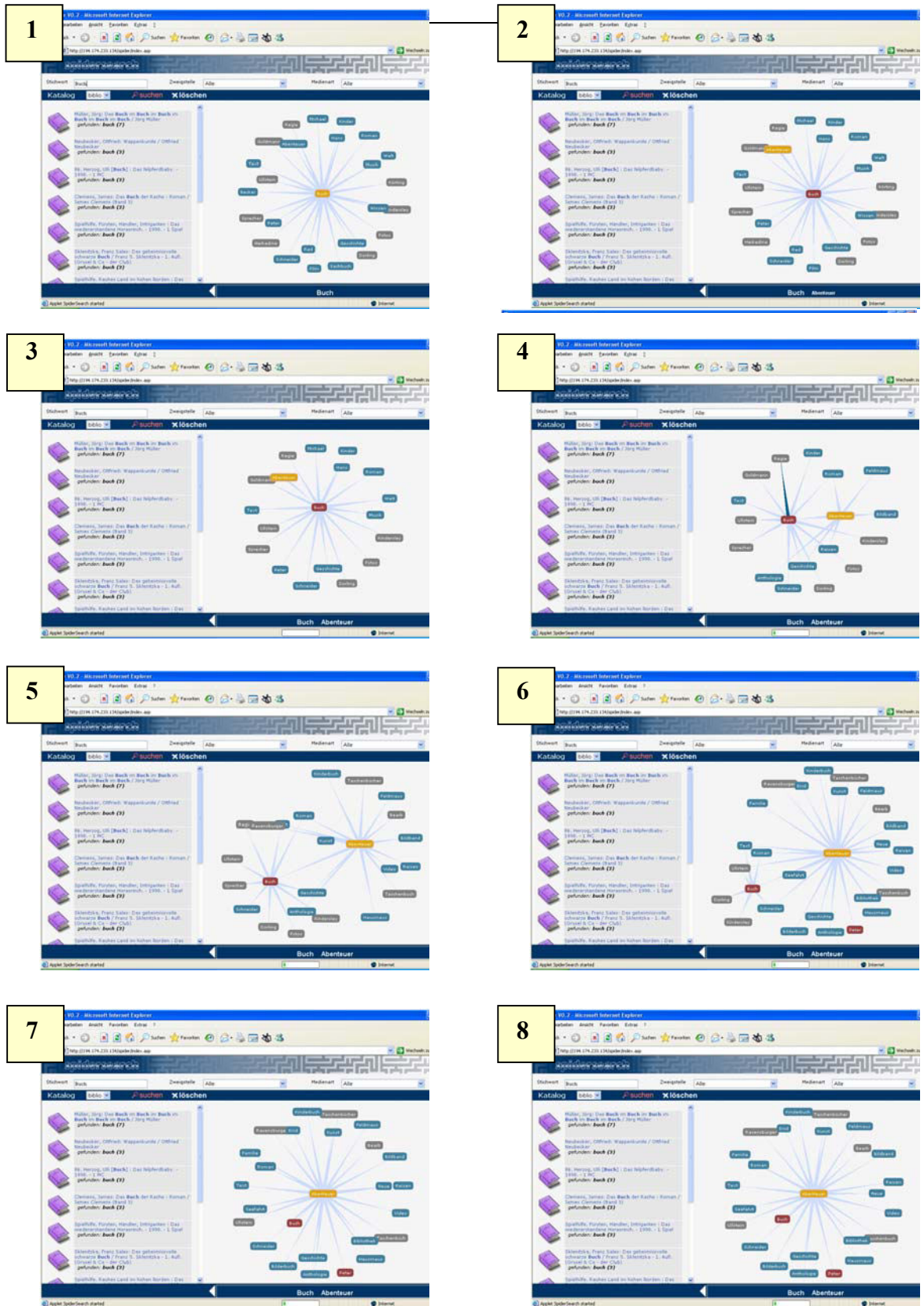


Abb. 35 N.N. 2005o – Animation des Netzübergangs von „Buch“ zu „Abenteuer“

Durch die Übernahme eines Begriffs aus dem semantischen Netz kann eine Suche weiterentwickelt werden: Der neu ausgewählte Begriff bildet in Kombination mit dem ursprünglichen Begriff die neue Suchanfrage. Bei einer Navigationsfolge durch das Netz werden jedoch nie mehr als zwei Begriffe kombiniert. Das heißt, dass jeweils der Term des aktuellen Netzes und der neu ausgewählte Begriff die neue Suchanfrage im Sinne einer booleschen UND-Verknüpfung bilden. Ausgehend von einer Suche nach „Buchdruck“ könnten die Navigationsschritte beispielsweise wie folgt aussehen:

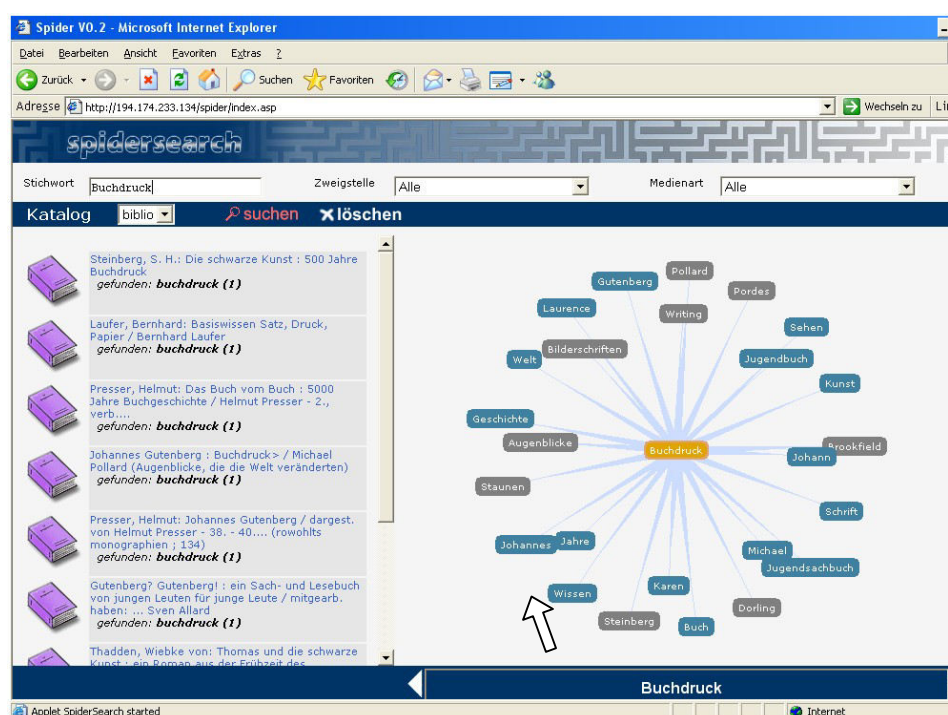


Abb. 36 N.N. 2005p – Suche nach „Buchdruck“, Auswahl des Terms „Wissen“

Über das Angebot im semantischen Netz kann der Begriff „Wissen“ in die Suche übernommen werden und bildet dann das Zentrum des neuen Netzes. Die Ergebnismenge wird aus der Verknüpfung des ersten Suchbegriffs („Buchwesen“) mit dem ausgewählten Begriff („Wissen“) erzeugt. Beide Suchwörter müssen in der Titelaufnahme¹⁹² enthalten sein, um zu einem Treffer zu führen. Die Sortierung nach Relevanz richtet sich nach der

¹⁹² Die Titelaufnahmen im Katalog der Stadtbibliothek Herten beinhalten neben formalen Angaben inhaltliche Erschließungsdaten wie Schlagwörter, Klassifizierung (Notation und natürlichsprachig) und Interessenkreise, die häufig um kurze Inhaltsangaben und Auszüge aus Klappentexten ergänzt sind. Die natürlichsprachigen Angaben der Systematik und die Interessenkreisangaben werden für die Ermittlung der Treffermenge nicht genutzt.

einfachen Termhäufigkeit, unabhängig von der Position eines Terms im Dokument. Die Angaben, wie häufig der Suchterm im gefundenen Dokument vorkommt, werden in der Trefferliste in Klammern jeweils aufgeführt:

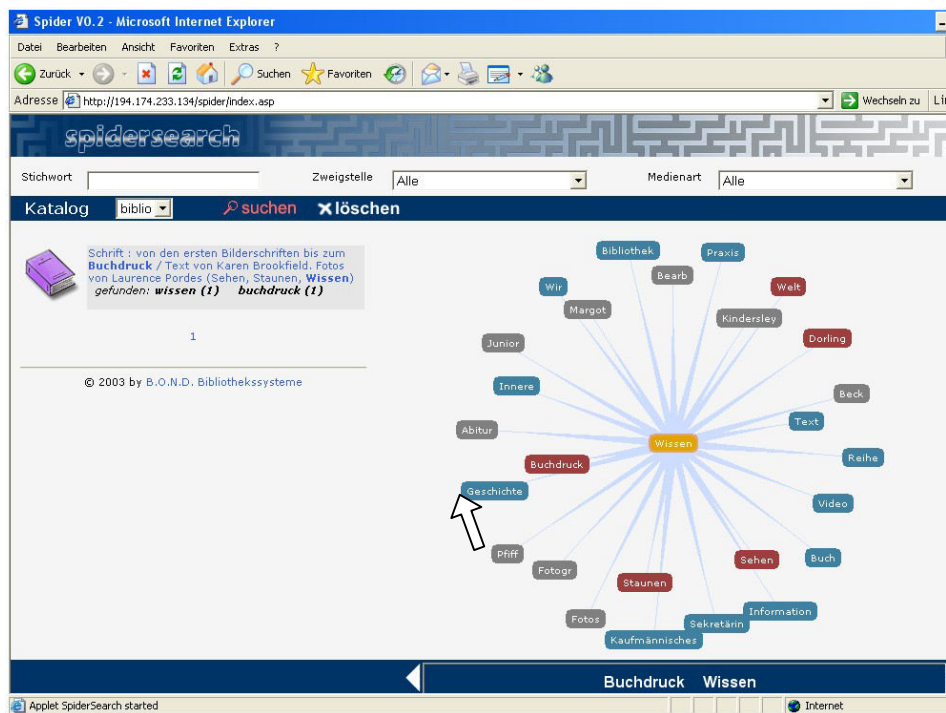


Abb. 37 N.N. 2005q – Auswahl des Eintrags „Geschichte“ im Netz zu „Wissen“

Aus dem neuen Netz des Terms „Wissen“ kann wiederum in einem weiteren Schritt ein neues Wort durch Anklicken in die Suche übernommen werden, wie beispielsweise der Begriff „Geschichte“. Für die Bildung der Treffermenge entfällt der ursprüngliche Suchbegriff „Buchdruck“. Die Ergebnismenge wird aus der UND-Verknüpfung der beiden zuletzt ausgewählten Terme „Wissen“ und „Geschichte“ gebildet. Es ist weder möglich, alle Terme aus einer Navigationsfolge über mehr als zwei Begriffe für eine Suche zu sammeln, noch eine ODER-Verknüpfung der Terme zu veranlassen. Die neue Ergebnismenge sieht wie folgt aus:

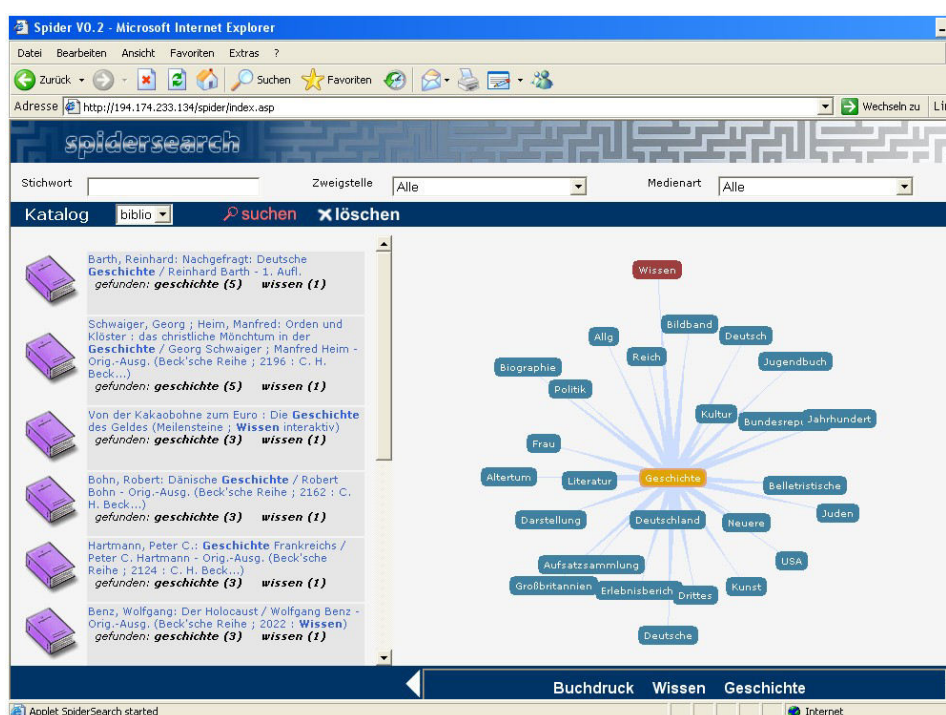


Abb. 38 N.N. 2005r – Netz zu „Geschichte“

Die Begriffsgenerierung des Netzes durch Eingabe eines Suchwortes unterscheidet sich nicht von der Begriffsgenerierung durch Termauswahl per Maus. Zu einem Begriff im Zentrum wird immer das gleiche Netz gebildet.

Eine Navigation durch die bereits erzeugten Netze ist über eine Dialogs-historie möglich. Durch Anklicken der Terme in der Begriffssammelleiste am unteren Bildschirmrand können bereits generierte Netze wieder aufgerufen werden. Die Pfeiltasten links und rechts dienen dem Scrollen in der Leiste, falls so viele Begriffe enthalten sind, dass sie aus Platzgründen nicht alle im sichtbaren Ausschnitt angezeigt werden können. Hier die Abbildung der Begriffssammelleiste:

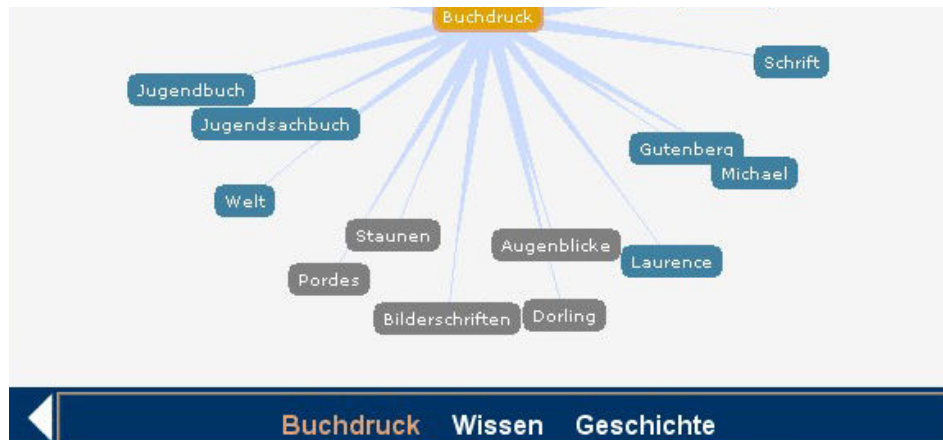


Abb. 39 N.N. 2005s – Begriffssammelleiste

Die Suche einer Phrase, wie beispielsweise „Öffentliche Bibliothek“, erzeugt das Netz des ersten eingegebenen Terms („Öffentliche“) und eine Treffermenge mit der Kombination beider Terme. Phrasen können nicht explizit als Phrasen gesucht werden, sondern die beiden Suchterme bilden eine UND-verknüpfte Anfrage, da Phrasen nicht indexiert und nicht im Netz ausgewiesen werden.

Folgende Darstellung wird zur Suche nach „Öffentliche Bibliothek“ erzeugt:

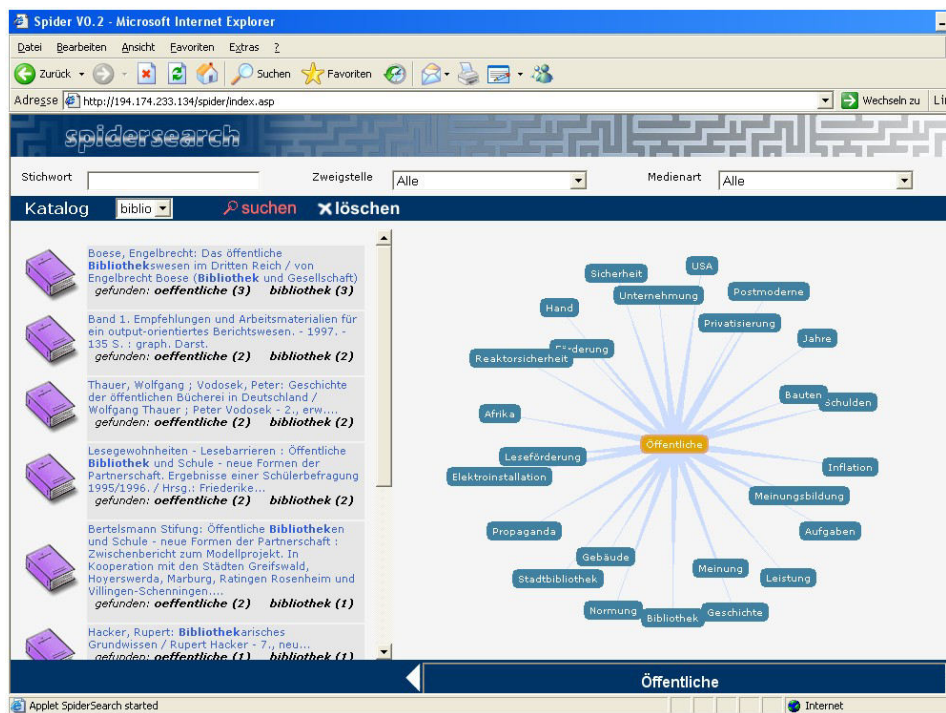


Abb. 40 N.N. 2005t – Suche nach „Öffentliche Bibliothek“¹⁹³

¹⁹³ Der erste Titel der Treffermenge „Das öffentliche Bibliothekswesen im Dritten Reich“ ist nicht auf eine automatisch trunkierte Suche zurückzuführen, sondern die Betrachtung der Titelaufnahme zeigt, dass die Terme „Bibliothek“ und „Öffentliche“ jeweils in der Reihenangabe und zweimal in den Schlagwortketten vorkommen und darüber zum Treffer führen. Der Term „Bibliothekswesen“ trägt nicht zur Trefferbildung bei.

Laut der Broschüre zu *SpiderSearch* soll der Nutzer das gewünschte Thema eines Homonyms mit „wenigen Klicks ... genau eingekreist“¹⁹⁴ haben. Als Beispiel wird die Suche nach „Käfer“ angegeben, zu der „sowohl Themengebiete wie Insekten als auch VW oder Wolfsburg zur Auswahl“¹⁹⁵ stehen sollen. Da die Terme des Begriffsnetzes jedoch abhängig vom Bestand der Anwenderbibliothek sind, gilt dies nicht in gleicher Wortwahl für die Stadtbibliothek Herten. Dort sind zwar sowohl Titel zum Insekt als auch zum Auto vorhanden, aber die erwähnten Begriffe „VW“ oder „Wolfsburg“ werden daraus nicht für das Netz gewonnen:

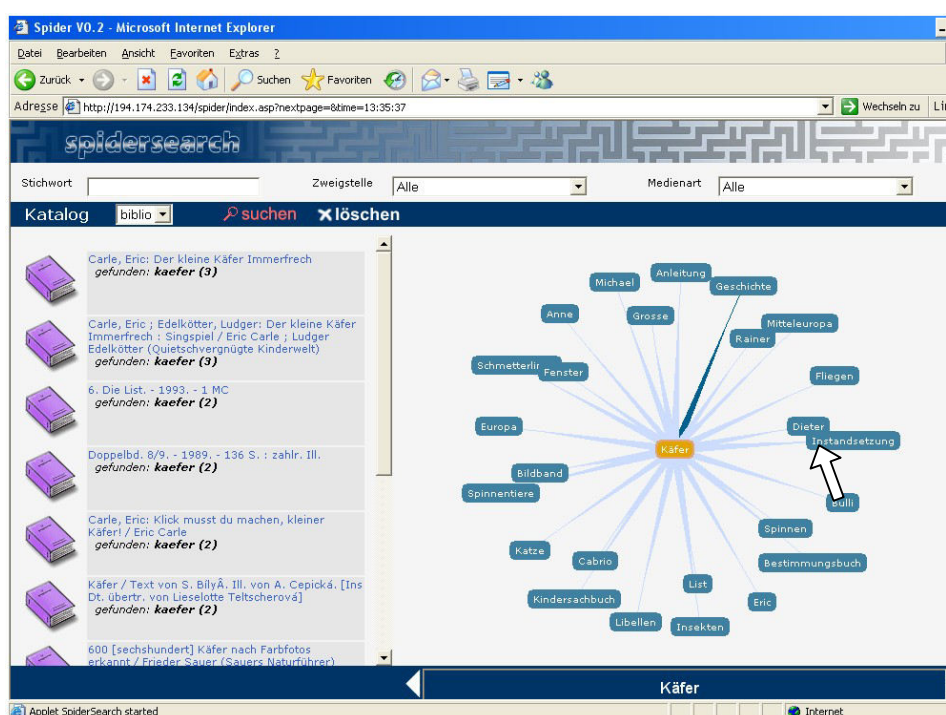


Abb. 41 N.N. 2005u – Auswahl des Eintrags „Instandsetzung“ im Netz zu „Käfer“

¹⁹⁴ „Broschüre zu SpiderSearch der Firma BOND“, S. 2

¹⁹⁵ Ebd.

Werden Titel zum Auto gewünscht, kann eine Einschränkung auf dieses Themengebiet im hier erzeugten Netz durch Auswahl des Terms „Instandsetzung“ oder „Cabrio“ erfolgen. Allerdings wird damit die Suche auch auf das zwingende Vorkommen der beiden Suchbegriffe „Käfer“ und „Instandsetzung“ beziehungsweise „Käfer“ und „Cabrio“ festgelegt und mögliche andere Titel zum Thema, die anderes Vokabular enthalten, werden von der Treffermenge ausgeschlossen. Bei einer Auswahl des Begriffs „Instandsetzung“ im Netz zu „Käfer“ wird folgender Titel ermittelt:

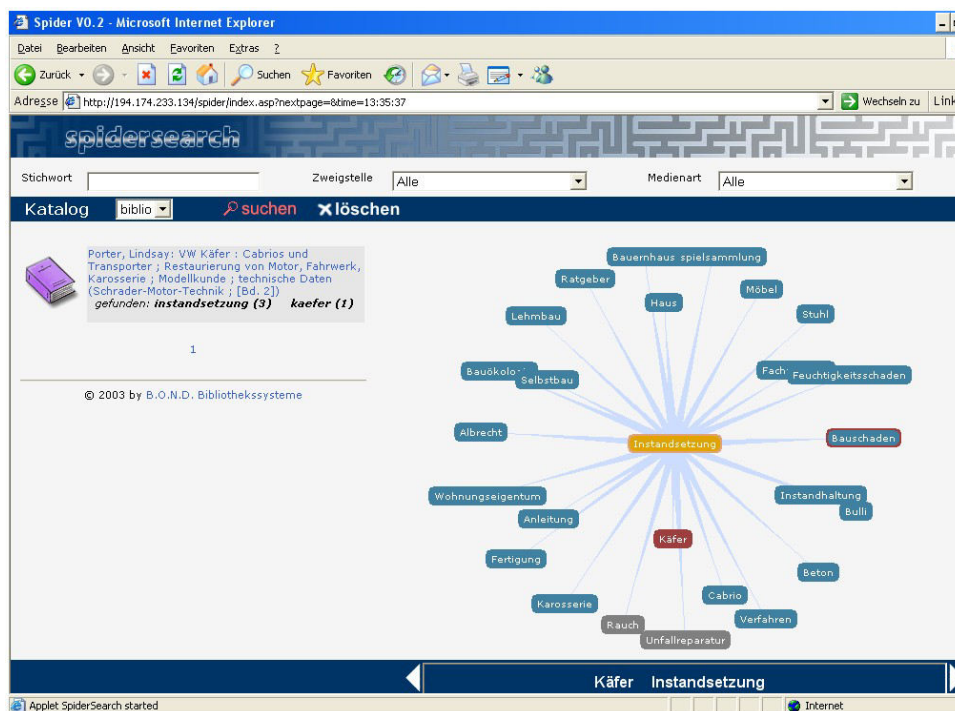


Abb. 42 N.N. 2005v – Netz zu „Instandsetzung“

Im Vergleich zu dem Versuch, die Suche auf diese Weise zu disambiguieren, können mit einer Suche nach dem sich eindeutig auf das Auto beziehende Schlagwort „VW-Käfer“ immerhin fünf Treffer ermittelt werden. Dieser Suchterm wurde jedoch nicht über das Netz angeboten, sondern aus der Schlagwortangabe des gefundenen Titels aus der oben genannten Suche übernommen. Zu einer Suche nach „VW-Käfer“ kann aus nicht ersichtlichen Gründen kein Netz angezeigt werden:

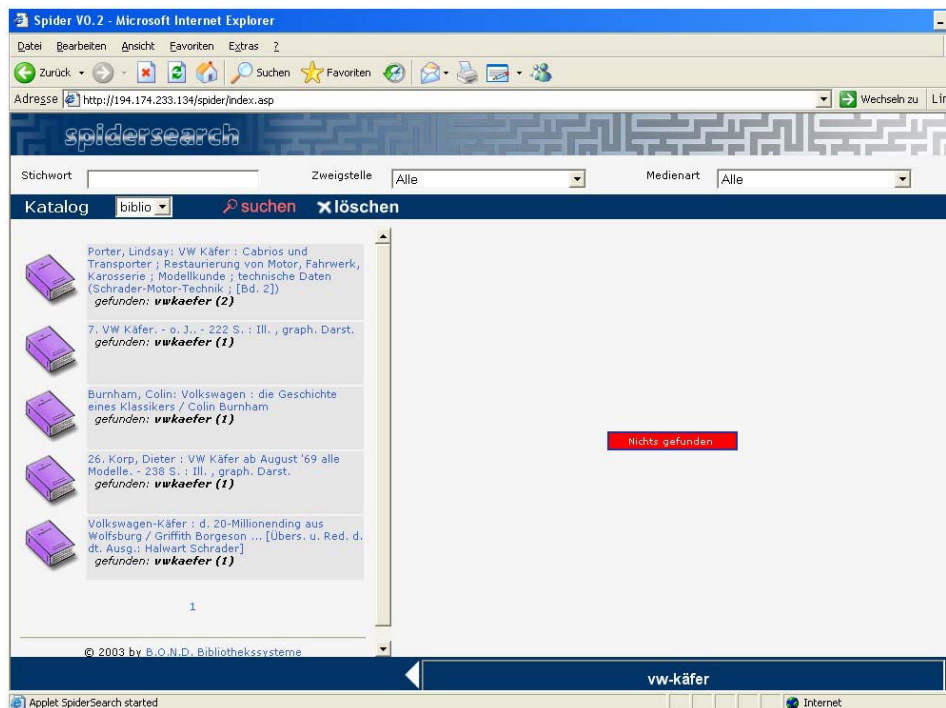


Abb. 43 N.N. 2005w – Suche nach „VW-Käfer“

Analog zur Bibliotheksanwendung *WordSurfer* ist es auch bei *Spider-Search* für das Ranking der Ergebnismenge problematisch, dass zwischen belletristischer Literatur und Sachliteratur nicht unterschieden wird und über die bibliographischen Titelaufnahmen sowohl auf formale als auch auf inhaltliche Aspekte gemischt zugegriffen wird. Es kann dadurch zu Treffern kommen, die vom Nutzer nicht gewünscht sind. Beispielsweise wird das formale Kriterium „Verfassernamen“ über die einfache Termhäufigkeit deshalb hoch gewichtet, weil es zweimal bei Verfasserwerken in der Titelaufnahme vorhanden ist. Eine Suche nach einem Personennamen, der gleichermaßen Sachbegriff ist, führt zu einer Treffermenge, in der Titel von gleichnamigen Verfassern vor Werken zum Thema gerankt sind, wie beispielsweise bei einer Suche nach „Rose“:

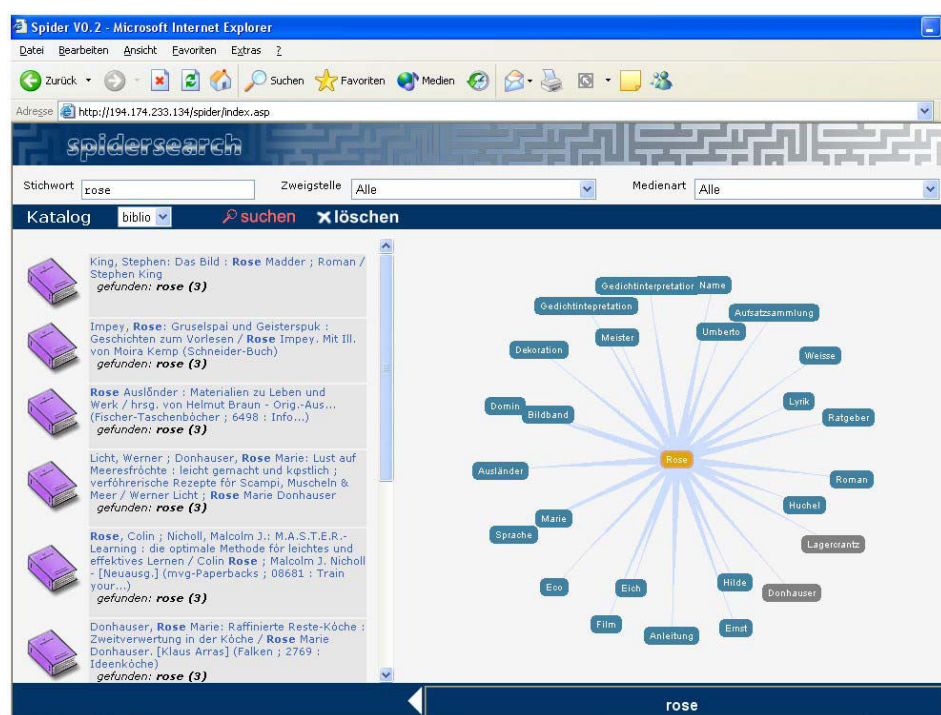


Abb. 44 N.N. 2005x – Suche nach „Rose“

Nach den Angaben auf der Homepage von *BOND* werden zu einer Suchanfrage verschiedene Schreibweisen eines Suchterms berücksichtigt. Als Beispiel wird eine Suche nach „Tschaikovsky“ aufgeführt, zu der auch alle Werke mit ähnlicher Schreibweise gefunden und die Namensvarianten im Netz ausgegeben werden sollen.¹⁹⁶ Bei einer Suche nach „Tschaikovsky“ in der Stadtbibliothek Herten wird nur ein Titel gefunden, hingegen bringt eine

¹⁹⁶ Vgl. „Homepage der Firma BOND – zu SpiderSearch“

Suche nach „Tschaikowsky“ 22 Treffer. Zu dieser Suchanfrage werden keine Namensvarianten im Netz ausgegeben. Eine Suche nach „Tchaikovsky“ führt zu 5 Treffern, eine Suche nach „Tschaikowski“ immerhin zu 20. Es ist also nicht unbedeutend, mit welcher Namensvariante die Suche formuliert wird. Die Trefferermittlung orientiert sich immer an der einfachen Termhäufigkeit durch Exact-Match zwischen Suchwort und Eintrag in der bibliographischen Angabe. Über die Auswahl von Namensvarianten, wenn sie denn ausgegeben werden, kann die Suche nur um den entsprechenden Term eingeschränkt werden. Eine Navigation durch das Netz dient also nicht der Erweiterung der Suchanfrage um unterschiedliche Schreibweisen von synonymen Benennungen.

5.4 Topic Maps Visualization in dandelon.com und Portal Informationswissenschaft

Die Firma *AGI – Information Management Consultants* bietet seit Frühjahr 2004 das Suchportal *dandelon.com*¹⁹⁷ an.¹⁹⁸ Es bietet über ein einheitliches Web-Interface Zugang zu den Katalogdaten verschiedener Bibliotheken an, wie beispielsweise zu den Titeln der Vorarlberger Landesbibliothek.¹⁹⁹ Das *Portal Informationswissenschaft*²⁰⁰, ein sich im Aufbau befindendes Portal, bietet Zugang zu bibliographischen Daten und zum Teil zu Volltexten von Publikationen der *Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis (DGI)* und zu Aufsätzen aus der Verbandszeitschrift *Information - Wissenschaft und Praxis (nfd)*.²⁰¹ Innerhalb dieser Portale können semantische Strukturen von intellektuell erstellten Thesauri visualisiert angezeigt und Begriffe für eine Suche ausgewählt werden.

Zur Anreicherung von bibliographischen Titelaufnahmen der beteiligten Bibliotheken und zur Indexierung der Daten setzt *AGI – Information Management Consultants* die Software *intelligentCapture* ein. Die Anreicherung erfolgt über gescannte Inhaltsverzeichnisse, die nach einer OCR²⁰² automatisch linguistisch indexiert werden.²⁰³ Darüber werden auch Personen- und Firmennamen als solche erkannt und geographische Terme wie Städtenamen ihren Regionen, wie Bundesland oder Staat, zugeordnet.²⁰⁴ In einem weiteren Schritt werden die Daten statistisch indexiert, wodurch ein Ran-

¹⁹⁷ Zugang über <http://www.dandelon.com>

¹⁹⁸ Vgl. Hauer, Manfred: Silicon Valley Vorarlberg: Maschinelle Indexierung und semantisches Retrieval verbessert den Katalog der Vorarlberger Landesbibliothek. In: Mitteilungen der VÖB 57 (2004) 3/4, S. 33-38, hier: S. 35

¹⁹⁹ Vgl. „Homepage des Portals dandelon.com – Frequently Asked Questions“

²⁰⁰ Zugang über http://www.agi-imc.de/isearch/is_dgi.nsf

²⁰¹ Vgl. „Homepage der Firma AGI - Information Management Consultants“

²⁰² Abkürzung für Optical Character Recognition

²⁰³ Vgl. Hauer, Manfred: Durch Content-Ergänzung, maschinelle Indexierung und modernes Information Retrieval können Recherchen in Bibliothekskatalogen deutlich verbessert werden. In: ABI-Technik 24 (2004) 4, S. 262-268, hier S. 265

²⁰⁴ Vgl. Hauer, Manfred: Digitalisierung von Aufsätzen und anderen Texten mit maschineller Inhaltserschließung erschließt Wissensressourcen in Bibliotheken, Museen und Dokumentationszentren, S. 8

king der Ergebnismenge zu einer Suchanfrage ermöglicht wird. Für das Retrieval wird neben der Software *intelligentCapture* zusätzlich die Software *intelligentSearch* eingesetzt. Sie ist auch „die technische Basis von dandelon.com“²⁰⁵ und vom *Portal Informationswissenschaft*.

Es können folgende Retrievalfunktionen für eine Suche verwendet werden: boolesche Operatoren, Links- und Rechtstrunkierung sowie Maskierung und kategorienspezifisches Suchen. Des Weiteren kann eine Phrasensuche durchgeführt oder es können Abstandsoperatoren eingesetzt werden. Eine automatische Fuzzy-Suche fängt Rechtschreibfehler ab.²⁰⁶ Die Suchbegriffe können um Thesaurus-Relationen erweitert werden. Dazu stehen in *dandelon.com* ca. 415.000 Terme aus verschiedenen Fachthesauri zur Verfügung.²⁰⁷

IntelligentSearch (und somit *dandelon.com* und das *Portal Informationswissenschaft*) bietet drei Suchmodi an, die per Radio-Buttons ausgewählt werden können. Die Standardeinstellung ist auf „more precise“ eingestellt. Damit wird das Suchwort um alle Synonyme, wie sie aus den Thesauri ermittelt werden, ergänzt. Sollte über diese Suche zu wenig oder nicht das gewünschte gefunden werden, kann die Suchanfrage in zwei weiteren Stufen erweitert werden. Nach der Wahl des Suchmodus „medium precise“ werden Unterbegriffe aus den Thesauri mit in die Suche einbezogen und über die Wahl des Suchmodus „more recall“ wird zusätzlich eine Links-Rechts-Trunkierung des Suchwortes durchgeführt.²⁰⁸ Neben den automatischen Expansionsmöglichkeiten der Suchanfrage gibt es die Möglichkeit, die semantischen Strukturen der Thesauri visualisiert anzuzeigen, darin zu navigieren und gezielt Suchbegriffe in die Anfrage zu übernehmen sowie durch Browsing im Netz auf weitere Begriffszusammenhänge aufmerksam zu werden. Nach Eingabe eines Suchwortes muss dazu der Button „Topic Maps Visualization“ angeklickt und der gewünschte Thesaurus ausgewählt werden, in dessen semantischer Struktur der Suchbegriff angezeigt werden soll.

Um im Rahmen dieser Arbeit auf den Nutzen der *SWD* für das Retrieval hinzuweisen, wurde exemplarisch ein Ausschnitt aus dem Bereich Bibliotheks- und Archivwesen im *Portal Informationswissenschaft* implementiert.²⁰⁹ Zusätzlich wurde der lückenhaften Verweisungsstruktur der *SWD*

²⁰⁵ Hauer, Manfred: Neue Mathematik in Bibliotheken? 2 = 100. AGI öffnet semantische Search-Engine für wissenschaftliche Literatur: www.dandelon.com, S. 2

²⁰⁶ Vgl. „Suchhilfe zum Portal Informationswissenschaft und dandelon.com“ und Hauer, Manfred: Silicon Valley Vorarlberg, S. 37

²⁰⁷ Vgl. „Homepage des Portals dandelon.com“

²⁰⁸ Vgl. Rädler, Karl: In Bibliothekskatalogen „googlen“. Integration von Inhaltsverzeichnissen, Volltexten und WEB-Ressourcen in Bibliothekskataloge. In: Bibliotheksdienst 38 (2004) 7/8, S. 927-939, hier: S. 933

²⁰⁹ Dazu wurde von der Verfasserin die *SWD* im Format BasisPlus in IC INDEX, der Thesaurusverwaltungsoftware für die Produkte *intelligentCapture* und *intelligentSearch*, importiert und freundlicherweise von Herrn Manfred Hauer (Unternehmensgründer, Senior Consultant und Leiter der Produktentwicklung der Firma AGI – Information Management Consultants) über das Portal *Informationswissenschaft* zugänglich gemacht. Der Bereich Bibliotheks- und Archivwesen der *SWD* im BasisPlus-Format ist im Anhang hinterlegt. Siehe „Die Schlagwortnormdatei aus dem Bereich Bibliotheks- und Archivwesen im BasisPlus-Format“

ein optimierter Zustand gegenübergestellt, so dass ein verbessertes Navigationsangebot gezeigt werden kann.²¹⁰

Die Eingabe des Suchbegriffs erfolgt in der Eingabezeile auf der Portal-seite²¹¹:

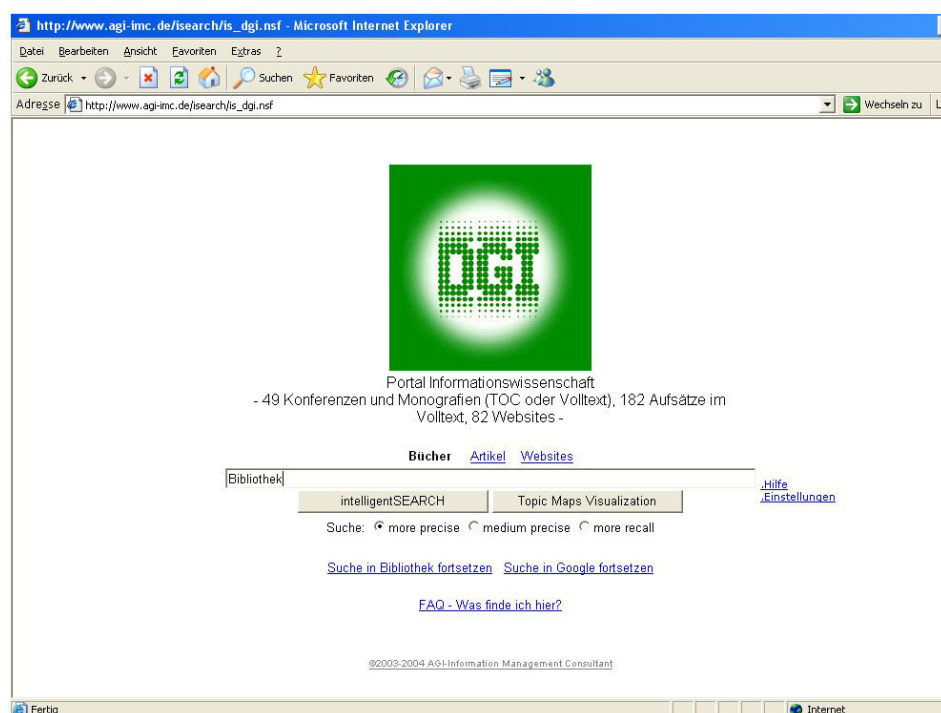


Abb. 45 Portal Informationswissenschaft – Suchoberfläche

Durch Drücken des Buttons „Topic Maps Visualization“ wird in einem ersten Schritt präsentiert, in welchen Thesauri der Suchbegriff vorhanden ist und wie viele Terme dort jeweils zum Suchbegriff relationiert sind. Die blauen Kreise, die die Termanzahl enthalten, sind umso dunkler, je mehr Terme zum Suchbegriff relationiert sind. Zur Suche nach „Bibliothek“ wird auf folgende Thesauri verwiesen:

²¹⁰ Der Bereich Bibliotheks- und Archivwesen der SWD wurde von der Verfasserin im Rahmen des in Kapitel 2.4.2 genannten Projekts optimiert, das heißt es wurden Doppelhierarchien und wenig oder inkonsistent hierarchisch relationierte Begriffe und wenig relationierte Strukturen ermittelt, verbessert und ergänzt. Der Ausschnitt der SWD ist im Portal Informationswissenschaft unter „Experimente mit SWD“ einzusehen, die optimierte Version unter „Experimente 2 mit SWD“.

²¹¹ Diese und folgende Angaben zur Suche im Portal Informationswissenschaft beruhen auf Erkundungen durch die Verfasserin. Die Angaben und Screenshots stützen sich auf den Online-Zugang über die Homepage der Firma AGI - Information Management Consultants. Siehe dazu „Homepage Portal Informationswissenschaft“, Zugang über http://www.agi-imc.de/isearch/is_dgi.nsf

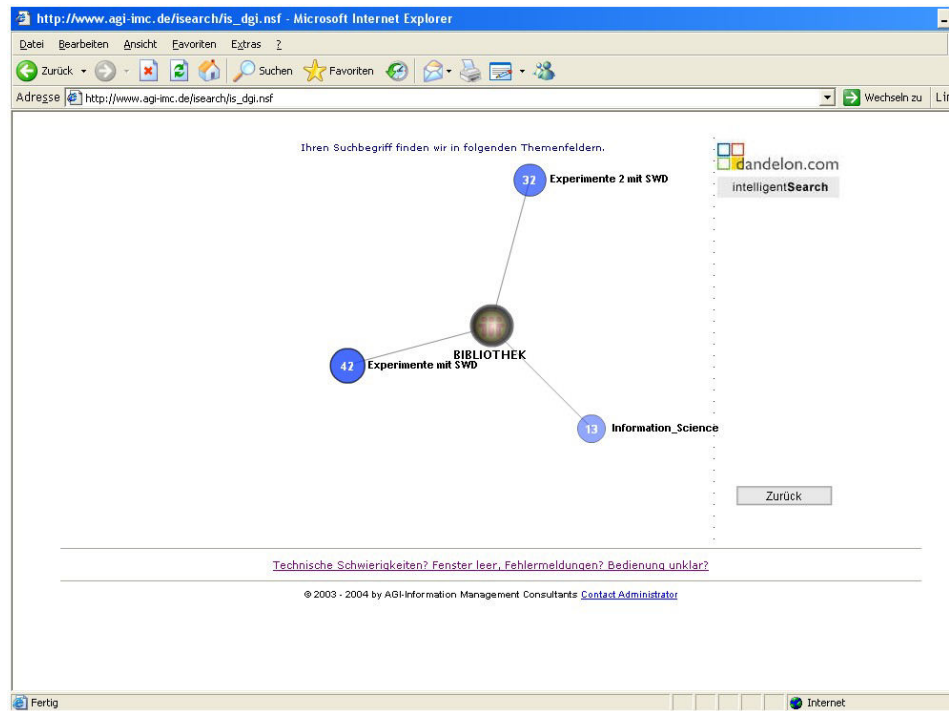


Abb. 46 Portal Informationswissenschaft – Thesaurusauswahl

Die blauen Kreise sind mit dem zugehörigen Thesaurus verlinkt. Durch Anfahren mit der Maus öffnet sich ein Pop-up-Fenster mit Detailangaben zum Thesaurus, wie Ersteller, Titel, Themenbereich, Sprache und Termanzahl. Durch Auswahl eines blauen Kreises wird die grafische Visualisierung des Suchbegriffs und seiner Relationierungen zu anderen Termen aufgerufen. Beispielsweise wird zu einer Suche nach „Bibliothek“ folgende Relationenstruktur angezeigt:²¹²

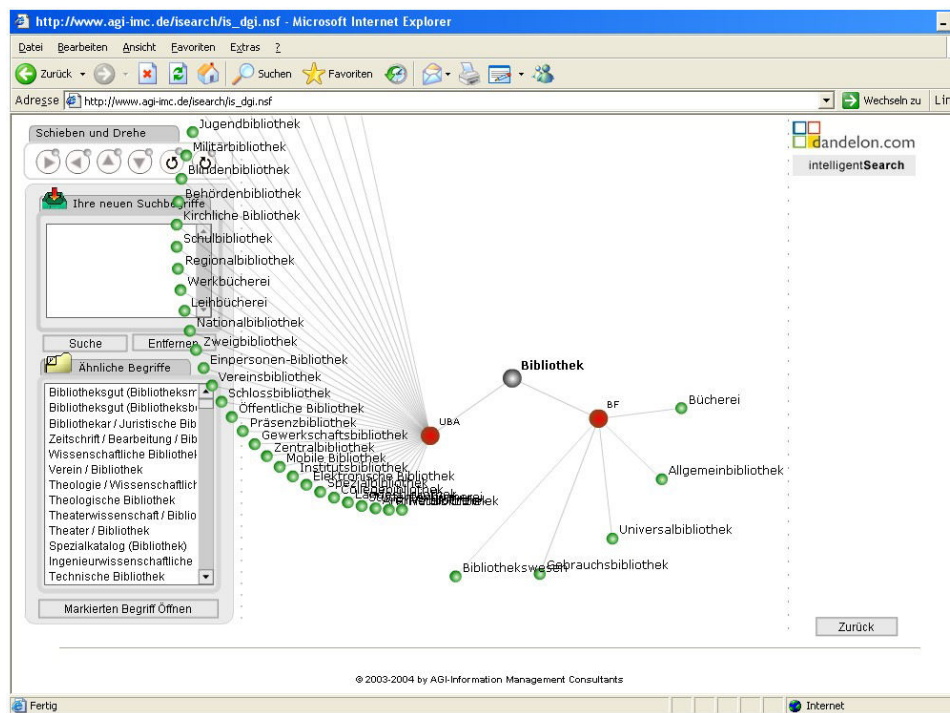


Abb. 47 Portal Informationswissenschaft – Topic Maps Visualization „Bibliothek“

²¹² Soweit nicht anders gekennzeichnet, wird in den folgenden Beispielen die Relationenstruktur der „Experimente mit SWD“ wiedergegeben.

Durch Interaktion kann der Thesaurusausschnitt gedreht und verschoben werden, um bei Überschneidungen von Termen, die dadurch nicht richtig gelesen werden können, die Perspektive auf die Begriffe nach Wunsch zu ändern. Über die Navigationsleiste „Schieben und Drehe [!]“ kann durch Anfahren der Knöpfe mit der Maus das Netz nach rechts, nach links, nach oben, nach unten oder rechts sowie links im Kreis gedreht werden. Hier die Abbildung der Navigationsleiste:



Abb. 48 Portal Informationswissenschaft – Navigationsleiste

Die Rotation erfolgt dabei relativ schnell, so dass es unter Umständen schwierig ist, die Begriffsstruktur genau an der Stelle anzuhalten, an der der gewünschte Begriff deutlich zu lesen ist. Dadurch, dass das Schieben und Drehen durch Mouse-Over und nicht durch Anklicken ausgelöst wird, kann leicht durch versehentliches Berühren eines Buttons eine Interaktion ausgelöst und das Netz ungewollt verändert werden.

Um den Suchbegriff im Zentrum der Darstellung werden die relationierten Begriffe angezeigt. Eine Interaktion mit dem Thesaurusausschnitt ist über die farbigen Knoten möglich. Dabei steht der graue Knoten für den Suchterm, auf den sich das aktuell aufgerufene Beziehungsgefüge bezieht. Die grünen Knoten dienen jeweils der Interaktion mit einzelnen Termen, die roten Knoten jeweils der mit einer Relationengruppe. Mittels der Knoten können entweder Terme aus dem Netz für eine Suche ausgewählt oder es kann durch die Netzstruktur navigiert werden. Eine Navigation durch den Thesaurus, also das Auswählen von Termen des aktuellen Netzes und das Aufrufen des jeweils zugehörigen Thesaurusausschnitts, erfolgt durch Anklicken der grünen Knoten. Der ursprüngliche Ausschnitt wird dabei durch das Begriffsnetz des neu ausgewählten Terms ersetzt. Eine Auswahl des grauen Knotens führt zu keiner Aktion, da der aktuell aufgerufene Thesaurusausschnitt bereits die Relationenstruktur dieses Terms darstellt. Eine Auswahl der Terme für eine Suchanfrage erfolgt durch Anfahren der Knoten mit der Maus.

Dadurch öffnet sich ein Pop-up-Fenster mit einer Schaltfläche „Zur Suche hinzufügen“, wie beispielsweise im Folgenden zum Netzeintrag „Arbeiterbibliothek“:

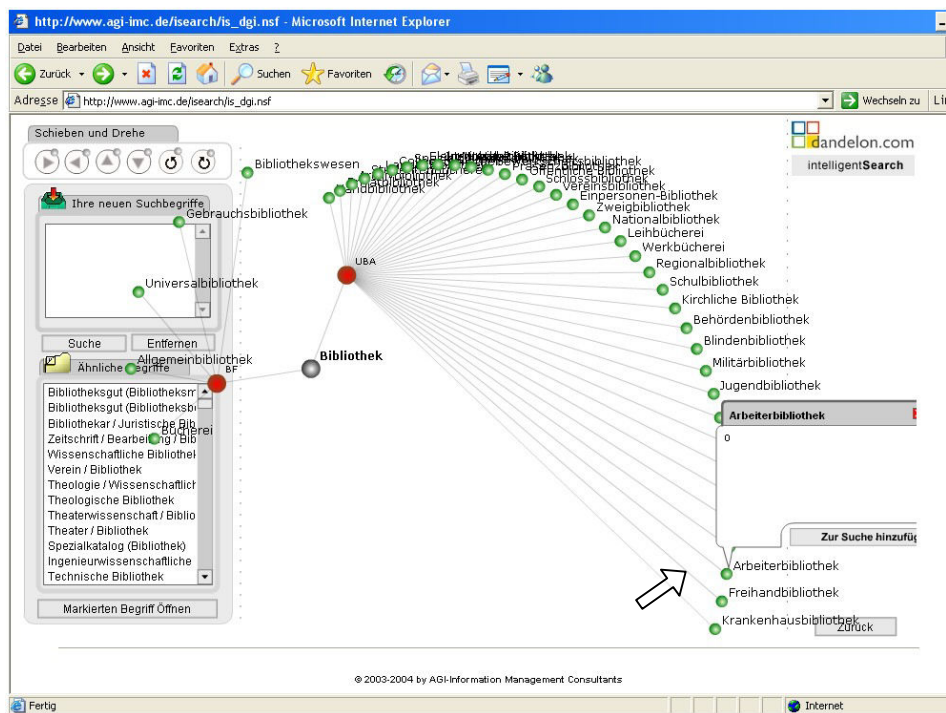


Abb. 49 Portal Informationswissenschaft – Pop-up-Fenster mit Schaltfläche „Zur Suche hinzufügen“

Durch Anklicken dieser Schaltfläche wird der ausgewählte Suchbegriff in die Sammelbox „Ihre neuen Suchbegriffe“ aufgenommen. Über die dort zusammengetragenen Terme kann eine neue Suchanfrage durch Anklicken des Buttons „Suche“ generiert werden. Die Terme aus der Box werden dann in die Eingabezeile des Portals übernommen und können für eine *intelligentSearch*-Suche genutzt werden. Sollen beispielsweise alle Unterbegriffe zu dem aktuell aufgerufenen Begriff „Bibliothek“ für eine Suche ausgewählt werden, können diese über den roten Knoten mit der Kennzeichnung „UBA“, die im Portal verwendete Abkürzung für Unterbegriff, auf einmal in die Sammelbox geschrieben werden. Die Auswahl der Begriffe kann nicht nur über das Pop-up-Fenster erfolgen, sondern alternativ auch über einen einfachen Klick mit der Maus auf den roten Knoten. Im Pop-up-Fenster werden Informationen zur Relationenart angegeben. Im folgenden Beispiel sind dies Angaben, die die Hierarchierelation erläutern. Als Relationenbenennung wird nur die Abstraktionsrelation („Abstrakte Hierarchie“) betitelt, obwohl in der *SWD* nicht zwischen generischen und partitiven Hie-

rarchierelationen unterschieden wird und beide gemischt als hierarchische Relationen ausgegeben werden:

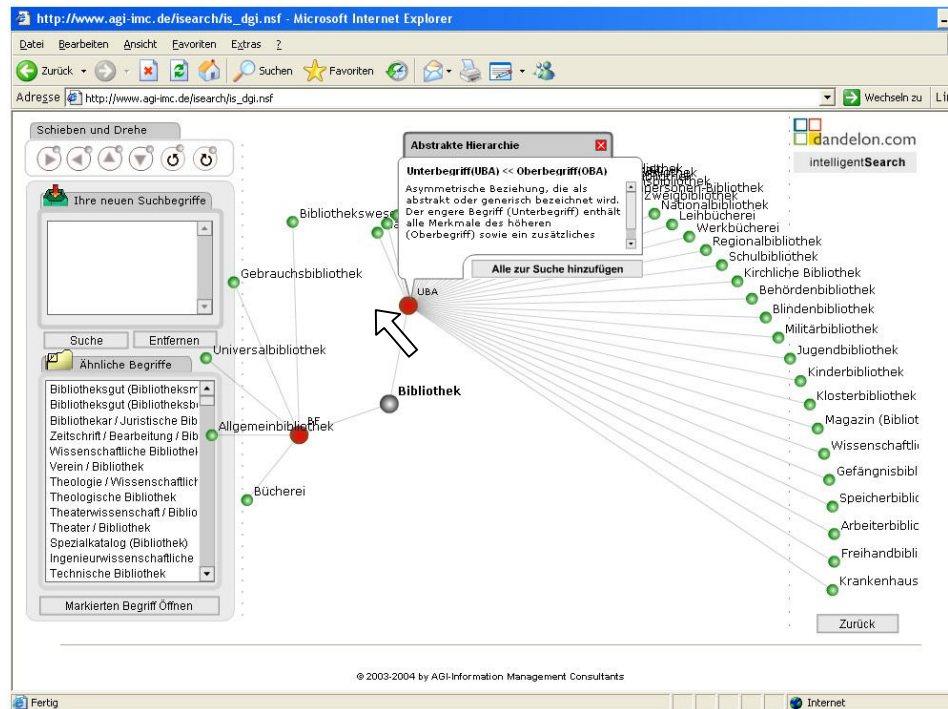


Abb. 50 Portal Informationswissenschaft – Pop-up-Fenster zur Hierarchierelation

Durch die Auswahl aller Unterbegriffe zum Schlagwort „Bibliothek“ sind für die Suchanfrage zu viele Begriffe in die Sammelbox geschrieben worden. Die Anzahl an verknüpften Begriffen einer Suchanfrage ist in *intelligentSearch* auf 19 Terme begrenzt; zum Schlagwort „Bibliothek“ liegen in der *SWD* 37 Unterbegriffe vor. Soll eine Suchanfrage zu den Unterbegriffen gestartet werden, müssen in einem weiteren Schritt also zunächst 18 Terme über den Button „Entfernen“ aus der Liste gelöscht werden. Erst dann können die verbleibenden 19 Terme eine neue Suchanfrage bilden. Durch die Übernahme der Terme in die Eingabezeile der Portaloberfläche findet automatisch eine ODER-Verknüpfung statt. Wird eine UND-Verknüpfung gewünscht, muss der boolesche Operator manuell überschrieben werden. Der ursprüngliche Suchterm „Bibliothek“ wird nicht wie erwartet für die Suche beibehalten. In der neuen Suchanfrage erscheint nur das, was auch vorher in die Sammelbox eingetragen wurde. Soll der Ursprungsterm mitgesucht werden, muss dieser Term explizit im Netz ausgewählt werden. Das geht jedoch nicht über den grauen Knoten des Suchterms im Zentrum, obwohl auch dort beim Anfahren mit der Maus das Pop-up-Fenster mit der Schaltfläche „Zur Suche hinzufügen“ aufklappt und die Möglichkeit der Begriffsübernahme

suggeriert. Tatsächlich kann der aktuelle Suchterm nur über einen Umweg in die Sammelbox geschrieben werden, indem beispielsweise ein Unterbegriff ausgewählt und der neue Thesaurusausschnitt zu diesem angezeigt wird. Dann kann der ursprüngliche Suchterm (hier: „Bibliothek“) als Oberbegriff (=grüner Knoten) ausgewählt werden. Alternativ könnte der Suchterm manuell durch Eintippen in der Eingabezeile ergänzt werden.

Die übersichtliche Darstellung der Thesaurusstruktur wird dadurch gestört, dass, bedingt durch das Anfahren von Knoten mit der Maus, ständig unbeabsichtigt Pop-up-Fenster geöffnet und dadurch große Teile des Netzes verdeckt werden. Sie können über ein rotes Kreuz in der rechten oberen Ecke zwar jeweils geschlossen werden, dies lenkt jedoch von der eigentlichen Nutzung des Vokabulars ab. Hinzu kommt, dass die Größe der Pop-up-Fenster statisch ist und sich nicht sensitiv zum Textinhalt vergrößert oder verkleinert. Die Fenster sind auch dann relativ groß und verdecken somit Vokabular, wenn sie gar keine Informationen zum Term enthalten.

Neben den hierarchischen und assoziativen Relationen werden ebenso die Äquivalenzrelationen zu einem Suchbegriff ausgegeben. Wird der Thesaurusausschnitt mit einem Synonym aufgerufen, erfolgt die Verweisung auf die Vorzugsbenennung, die mit dem roten Relationenknoten und mit „BS“ (Benutze Synonym) gekennzeichnet ist. Bei einer Suche mit einem Schlagwort werden die synonymen Begriffe mit „BF“ (Benutzt für) bezeichnet. Für die Orientierung und Navigation im Begriffsnetz kann es hinderlich sein, dass synonyme Einträge mit den gleichen Knotenfarben versehen sind wie Verweise auf Unterbegriffe, Oberbegriffe und verwandte Begriffe, denn die Farbgebung springt stärker ins Auge, als die schriftliche Relationenartkennzeichnung. Es könnte dadurch passieren, dass ein Nutzer synonyme Einträge nicht gleich als solche erkennt und bei der Navigation durch das Netz auswählen möchte. Dieser Schritt führt dann jedoch nicht zu weiteren Begriffen, sondern nur zu dem Verweis auf die bereits bekannte Vorzugsbenennung.

Im Folgenden wird dies durch die dargestellten Einträge zum Schlagwort „Bibliothek“ und der Auswahl des Synonyms „Allgemeinbibliothek“ aufgezeigt:

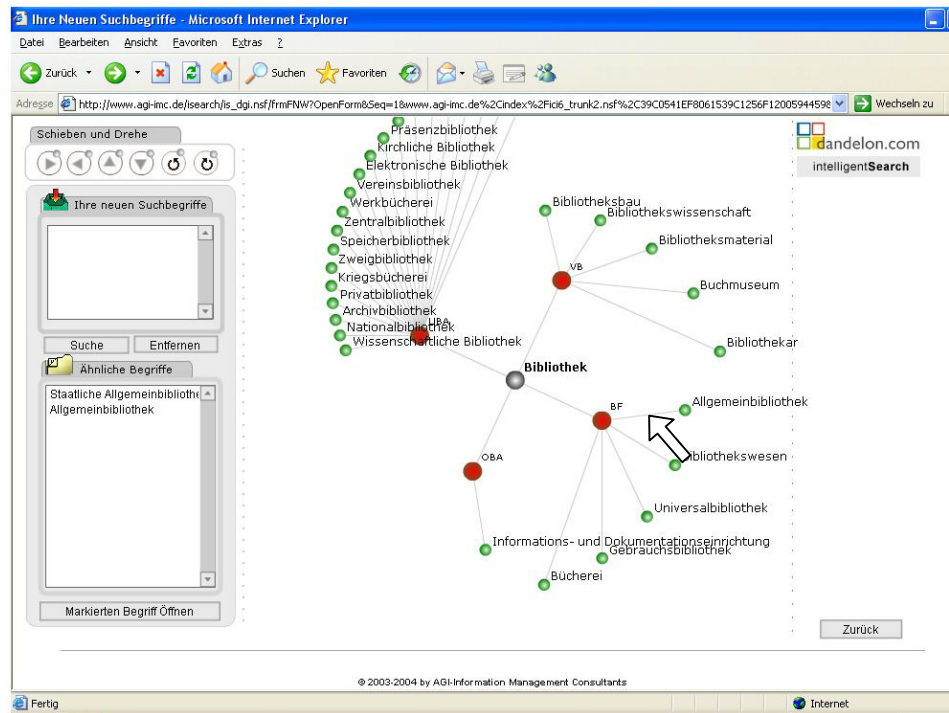


Abb. 51 Portal Informationswissenschaft – Auswahl des Eintrags „Allgemeinbibliothek“ im Netz zu „Bibliothek“

Nach Auswahl des Synonyms „Allgemeinbibliothek“ folgt die Darstellung mit der Verweisung zurück auf die Vorzugsbenennung „Bibliothek“:

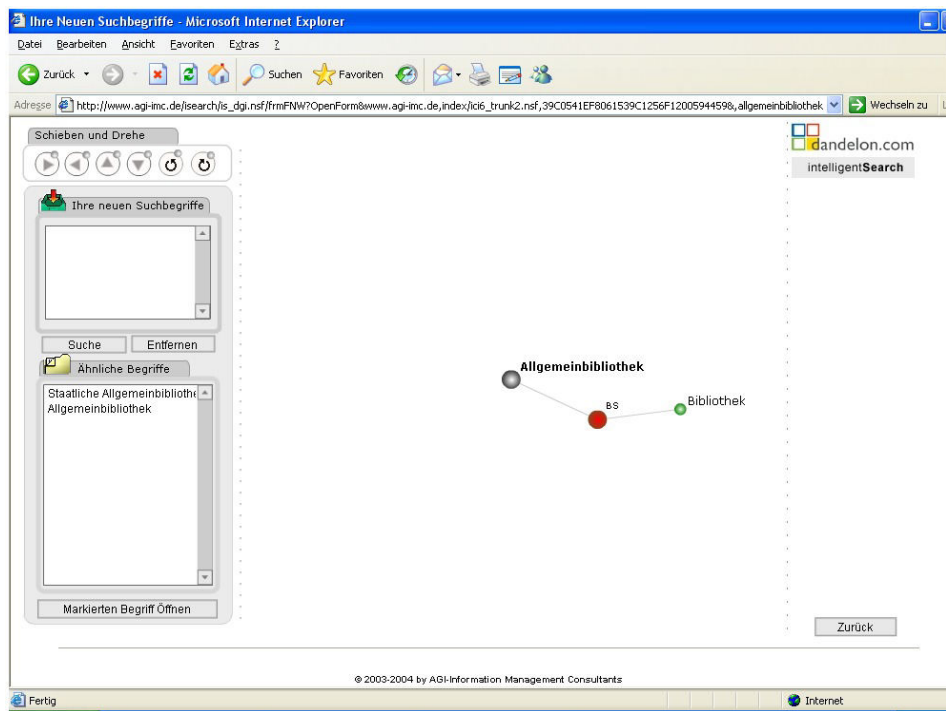


Abb. 52 Portal Informationswissenschaft – Topic Maps Visualization „Allgemeinbibliothek“

Eine Disambiguierung von Homonymen erfolgt in der *SWD* durch Homonymenzusätze. Beispielsweise liegen zu dem Begriff „Magazin“ die Einträge „Magazin (Archiv)“ und „Magazin (Bibliothek)“ vor. Eine Suche nach „Magazin“ im *Portal Informationswissenschaft* führt zur Anzeige des nach alphabetischer Reihenfolge ersten Eintrags. Hier der Eintrag mit dem Zusatz „Archiv“:

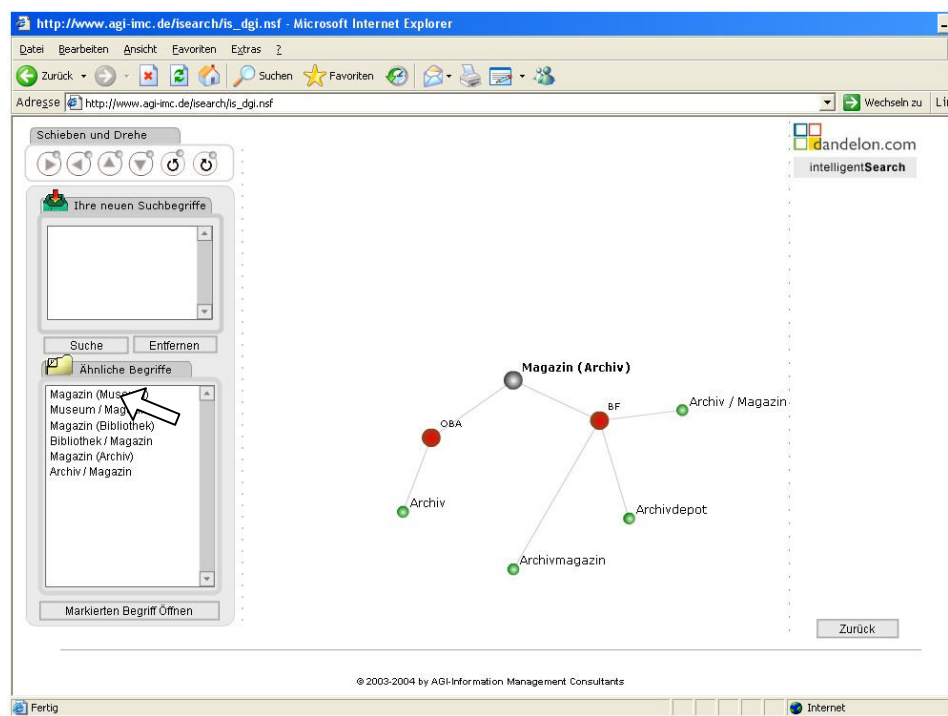


Abb. 53 Portal Informationswissenschaft – Topic Maps Visualization „Magazin (Archiv)“

Das Schlagwort „Magazin (Bibliothek)“ kann über die Sammelbox „Ähnliche Begriffe“²¹³ aufgerufen werden. Dazu muss der Eintrag mit der Maus ausgewählt und über den Button „Markierten Begriff Öffnen“ aufgerufen werden. In der Begriffssammlung „Ähnliche Begriffe“ wird zwischen Vorzugsbenennungen und Synonymen nicht unterschieden.

²¹³ Die Sammelbox „Ähnliche Begriffe“ enthält alle Schlagwörter, in denen der Suchterm vorkommt.

Die Suche nach einer Phrase führt zu einer Visualisierung des zusammengesetzten Begriffs, wenn ein entsprechender Eintrag im Thesaurus vorhanden ist. Beispielsweise gibt es zur Suche nach „Information Retrieval“ folgende Auswahlmöglichkeiten:

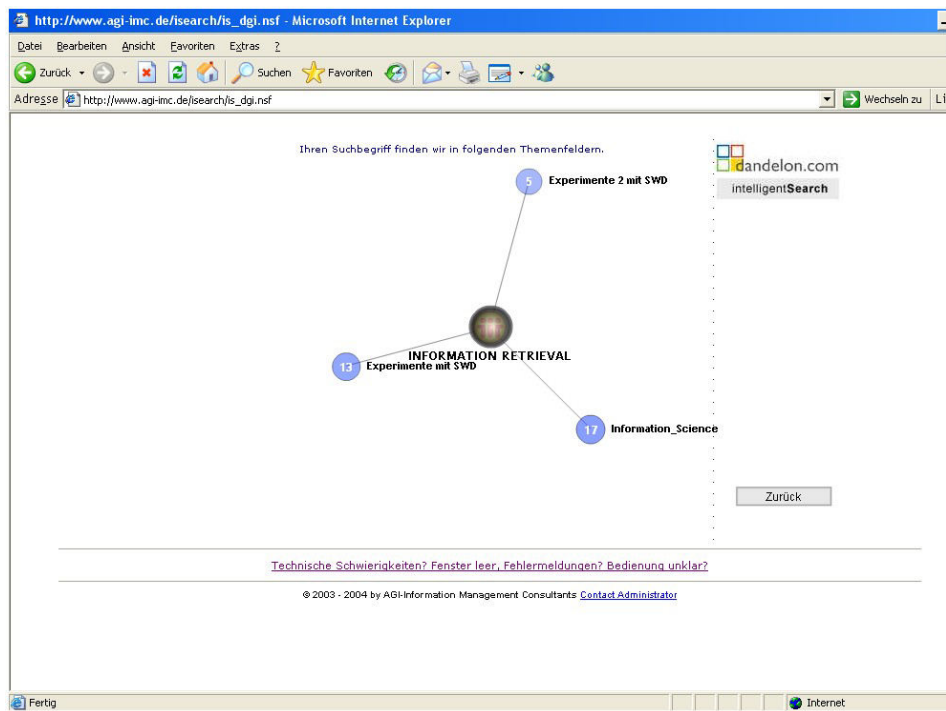


Abb. 54 Portal Informationswissenschaft – Thesaurusauswahl zu „Information Retrieval“

Nach Auswahl des Thesaurus „Experimente mit SWD“ wird das Relationengefüge zur Phrase „Information Retrieval“ visualisiert:

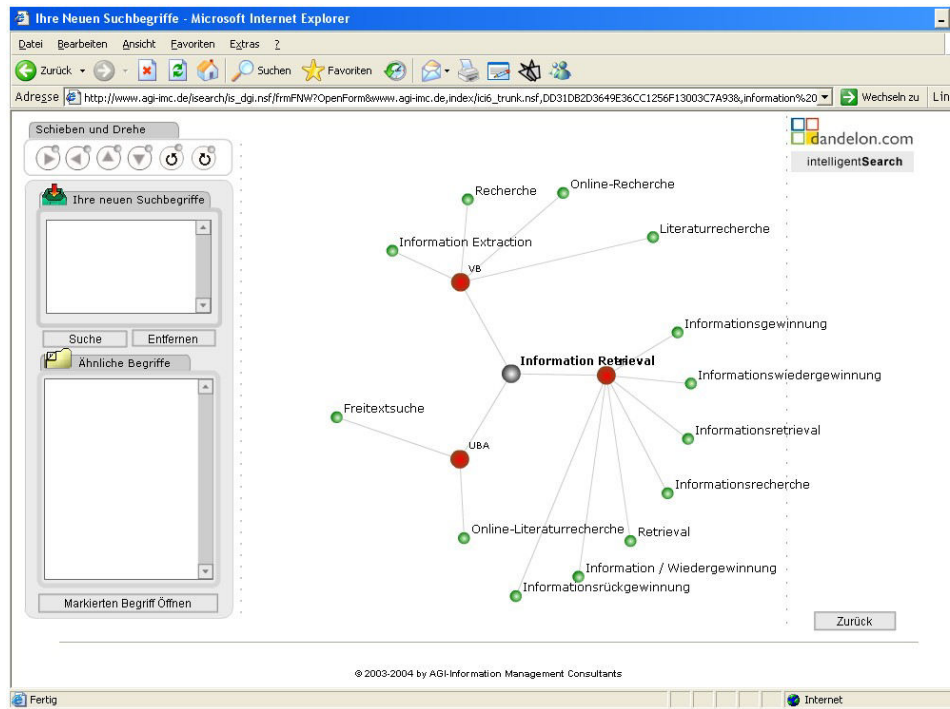


Abb. 55 Portal Informationswissenschaft – Topic Maps Visualization „Information Retrieval“

Kombinierte Begriffe, die keinem Phraseneintrag im Thesaurus entsprechen, können nicht angezeigt werden, da die Darstellung zweier Netze nicht möglich ist. Eine Suche beispielsweise nach „Blindenbibliothek Hörbibliothek“ führt zum Hinweis, dass dieser Suchbegriff in keinem Thesaurus gefunden werden konnte:

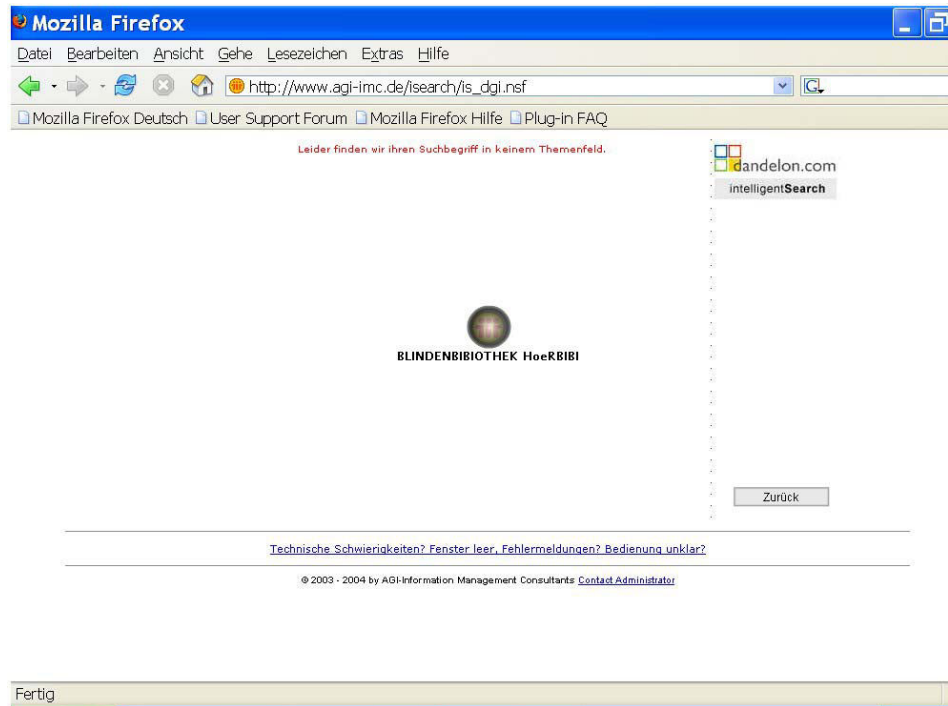


Abb. 56 Portal Informationswissenschaft – Thesaurauswahl zu „Blindenbibliothek Hörbibliothek“

Wird eine Begriffskombination verknüpft durch boolesche Operatoren gesucht, besteht theoretisch die Möglichkeit, mittels eines Pull-down-Menüs den Eintrag zu wählen, der zuerst angezeigt werden soll. Praktisch springt die Auswahl im Menü immer auf den ersten Term zurück, so dass zu einer Suche wie nach „Blindenbibliothek AND Hörbibliothek“ nur die Möglichkeit besteht, sich die Relationenstruktur zu „Blindenbibliothek“ anzeigen zu lassen. Hier die Anzeige der Auswahlbox:

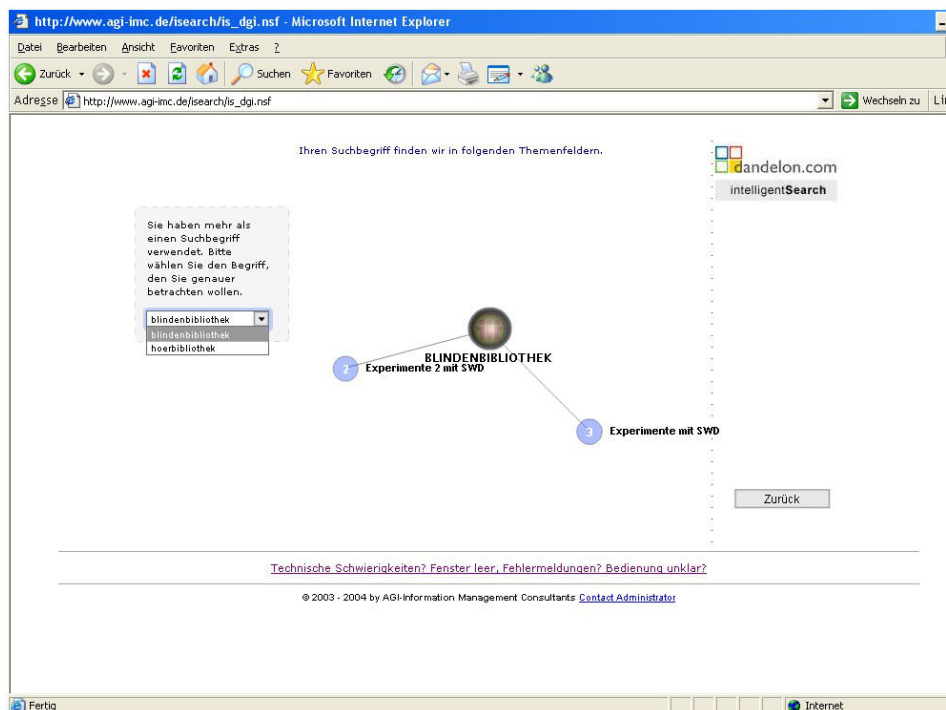


Abb. 57 Portal Informationswissenschaft – Thesaurusauswahl zu „Blindenbibliothek AND Hörbibliothek“

Im Folgenden werden drei Begriffsrelationen aus der *SWD* denen aus der nachrelationierten Version gegenübergestellt. Es soll exemplarisch gezeigt werden, wie sich lückenhafte Relationierungen negativ auf die Navigation im semantischen Netz auswirken können und wie eine optimierte Struktur den Retrievalprozess unterstützen kann. Beispielsweise ist das Schlagwort „Dezimalklassifikation“ in der *SWD* nicht relationiert. Im Vergleich dazu ist der Begriff in der nachrelationierten Version in eine hierarchische Struktur integriert und es kann dadurch zu weiteren Klassifikationstypen navigiert werden. Der Eintrag in der *SWD*:

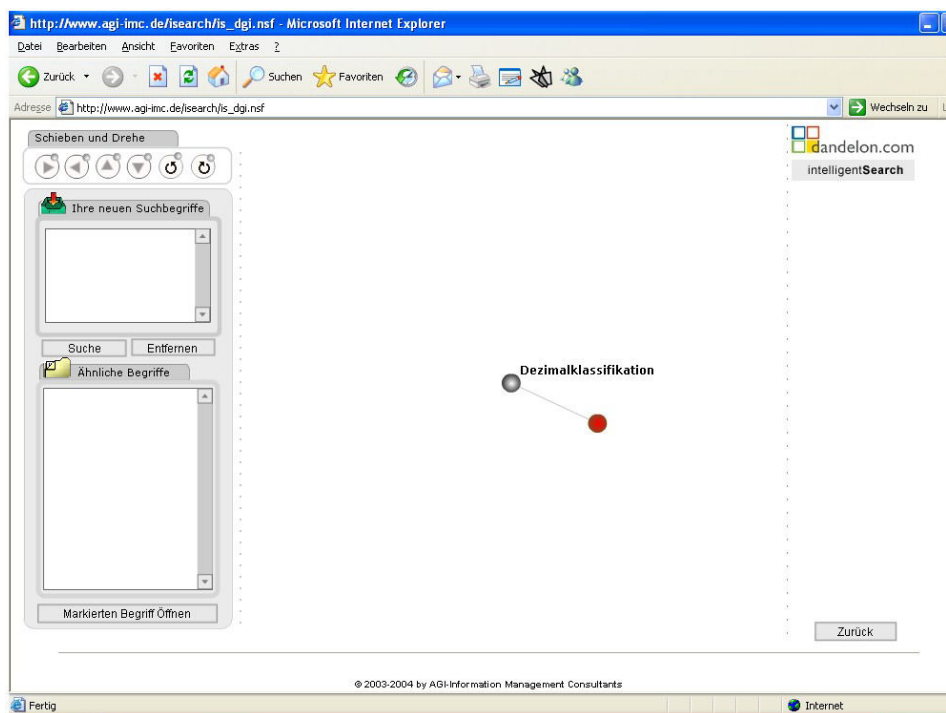


Abb. 58 Portal Informationswissenschaft – Topic Maps Visualization
„Dezimalklassifikation“ aus „Experimente mit SWD“

Zum Vergleich der Eintrag in der nachrelationierten *SWD*:

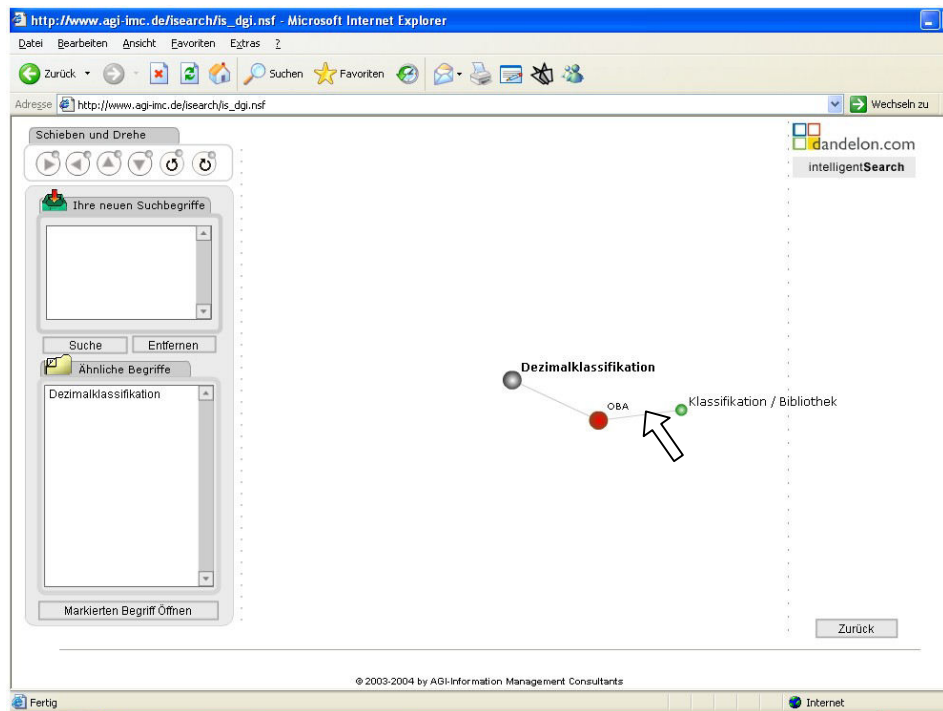


Abb. 59 Portal Informationswissenschaft – Topic Maps Visualization
„Dezimalklassifikation“ aus „Experimente 2 mit SWD“

Durch die hierarchische Relationierung kann im Netz navigiert und neue Begriffe können aufgerufen werden, wie beispielsweise hier „Klassifikation / Bibliothek“:

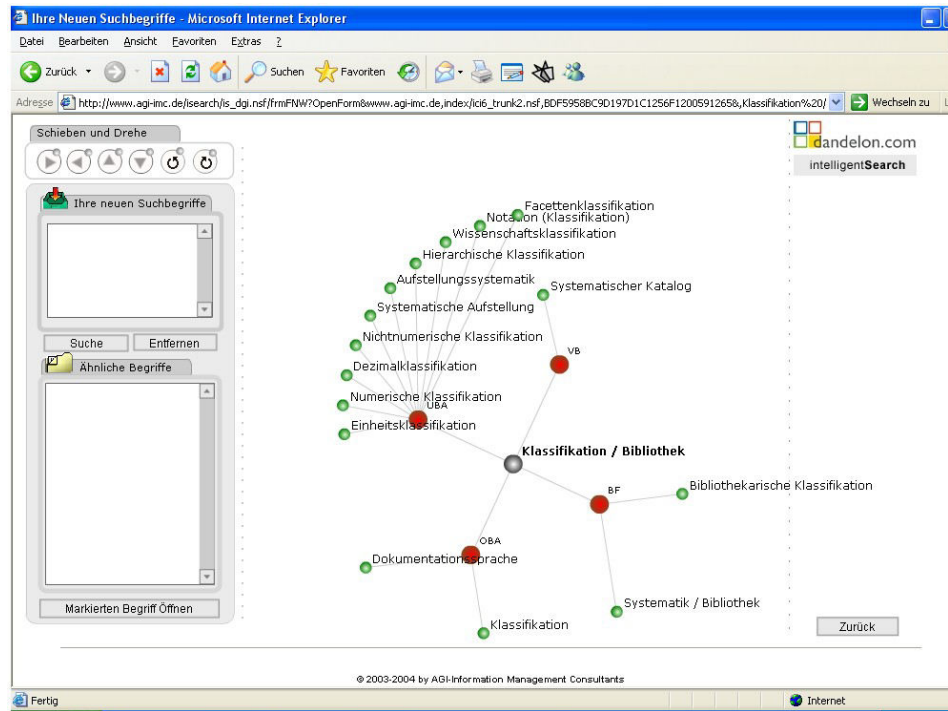


Abb. 60 Portal Informationswissenschaft – Topic Maps Visualization „Klassifikation / Bibliothek“ aus „Experimente 2 mit SWD“

Beispielsweise ist in der *SWD* das Schlagwort „Buchaufstellung“ ausschließlich zu synonymen Benennungen relationiert. Es fehlen Angaben zu Begriffen aus assoziativen und hierarchischen Beziehungen. Die Begriffe selbst sind in der *SWD* vorhanden, wurden aber erst durch die Nachrelationierung mit dem Schlagwort „Buchaufstellung“ verknüpft. Der Eintrag in der *SWD*:

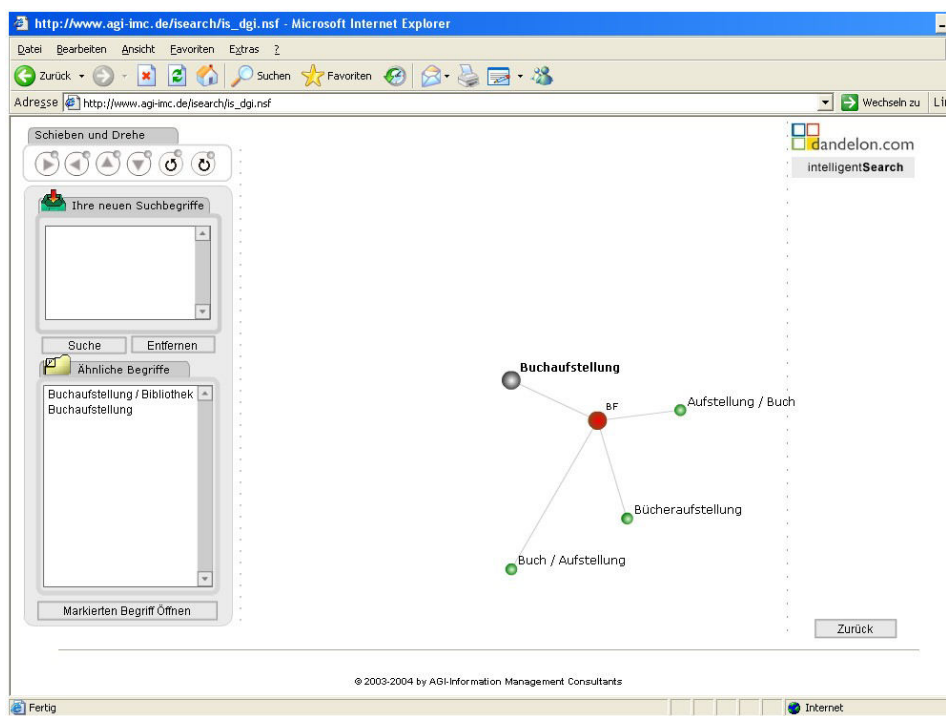


Abb. 61 Portal Informationswissenschaft – Topic Maps Visualization „Buchaufstellung“ aus „Experimente mit SWD“

Zum Vergleich der Eintrag in der nachrelationierten *SWD*:

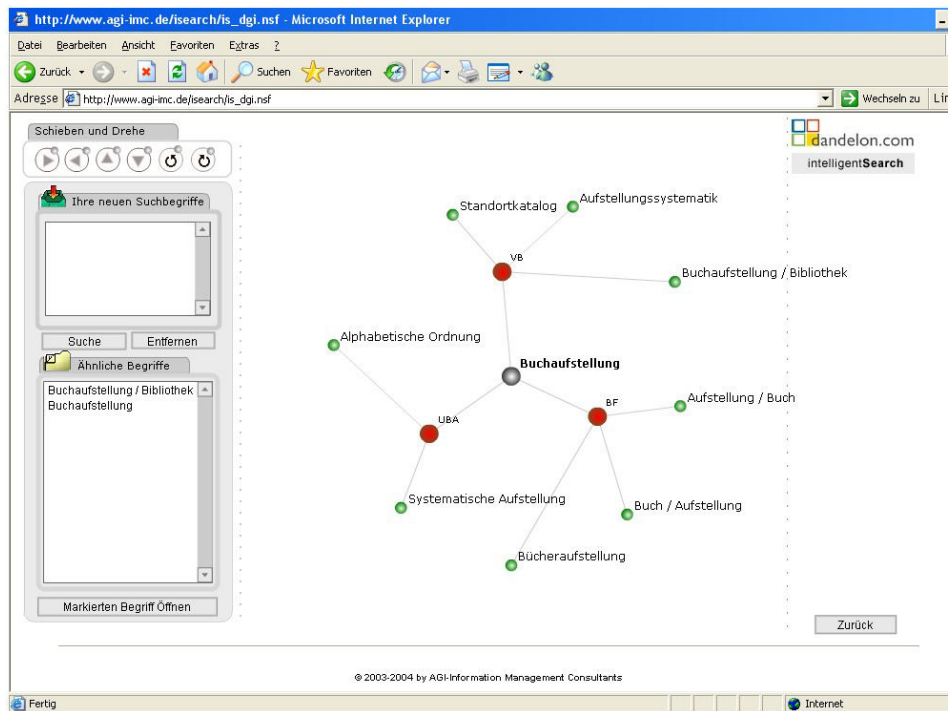


Abb. 62 Portal Informationswissenschaft – Topic Maps Visualization „Buchaufstellung“ aus „Experimente 2 mit SWD“

Beispielsweise ist in der *SWD* das Schlagwort „Thesaurus“ nur zu einem weiteren Begriff assoziativ relationiert. In der für Retrievalzwecke optimierten Version wurden weitere verwandte Begriffe ergänzt, so dass durch Interaktion mit dem Vokabular die Möglichkeit besteht, eine Suche um assoziative Zusammenhänge zu erweitern. Der Eintrag in der *SWD*:

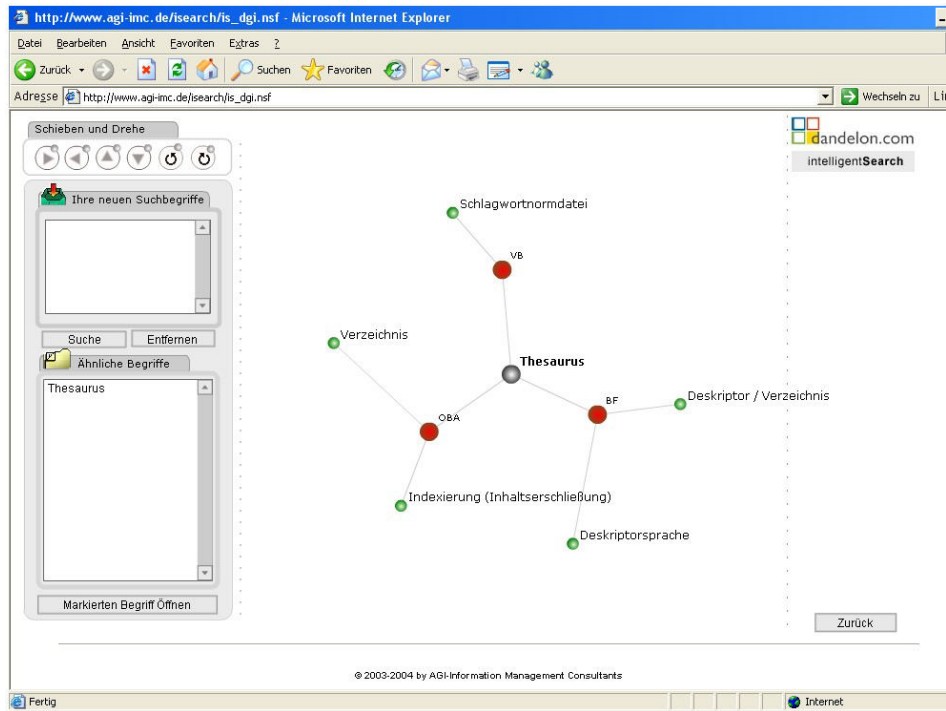


Abb. 63 Portal Informationswissenschaft – Topic Maps Visualization „Thesaurus“ aus „Experimente mit SWD“

Zum Vergleich der Eintrag in der nachrelationierten *SWD*:

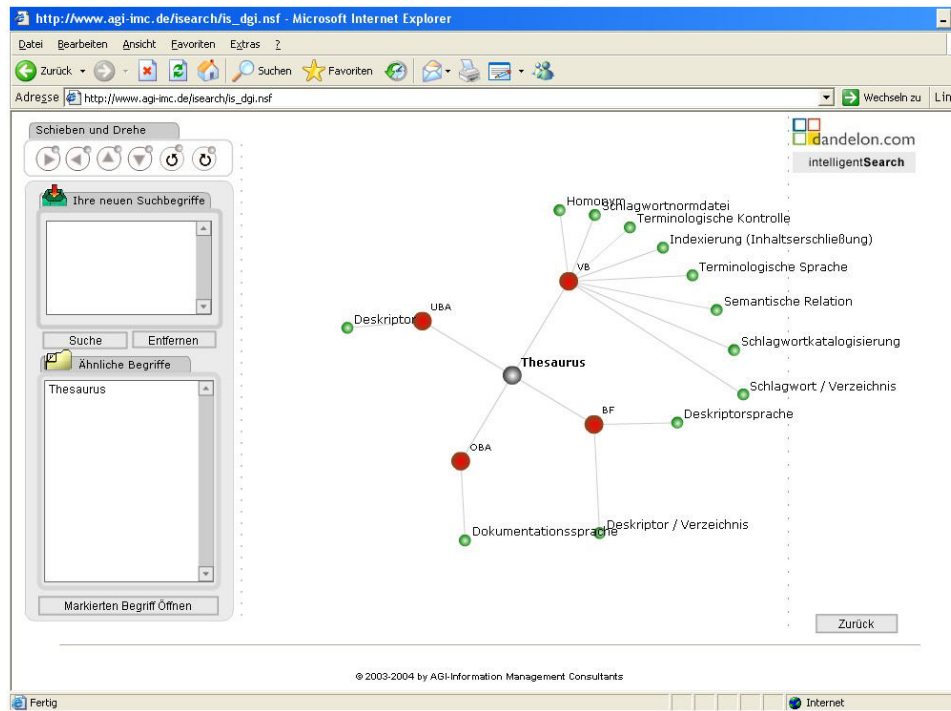


Abb. 64 Portal Informationswissenschaft – Topic Maps Visualization „Thesaurus“ aus „Experimente 2 mit SWD“

6 RESÜMEE UND AUSBLICK

In dieser Arbeit sollte gezeigt werden, dass durch Einsatz von semantischen Netzen und insbesondere durch deren grafische Darstellung in Kombination mit interaktiver Nutzungsmöglichkeit der Suchprozess verbessert werden kann.

Es wurden Anwendungen vorgestellt, die die Visualisierung von semantischen Netzen und die Interaktionsform der Hypertexttechnik zur Unterstützung des Retrievalprozesses einsetzen. Dabei konnte auch festgestellt werden, dass für die Verbesserung der Suche nicht nur die in dieser Arbeit behandelten Aspekte der Visualisierung und Interaktion entscheidend sind, sondern auch die Qualität der Begriffe und Relationen. Eine für statistische automatische Indexierungsverfahren nicht ausreichende Datenbasis führt nicht zu hilfreichen Netzeinträgen. Damit bietet sich vor allem für Bibliotheken neben einer Anreicherung der bibliographischen Daten die Nutzung der bereits für Erschließungszwecke vorhandenen Begriffsstruktur *SWD* an. Intellektuell erstelltes Vokabular mit sauber ausgearbeiteten Relationen kann statistisch ermittelte Begriffsrelationen um wichtige Einträge ergänzen. Für die *SWD* stellt sich dafür die Forderung nach einer Optimierung der Struktur im Hinblick auf ein Instrument zum Suchen und Finden.

Dass Visualisierung auch dazu beitragen kann, das Suchen in Informationssystemen für den Nutzer vergnüglich zu gestalten, ist ein weiterer Aspekt, der nicht übergangen werden sollte. So beschreiben Bederson u.a. das Ziel ihrer Arbeit im University of marylands's Human-Computer Interaction Lab, dessen Forschungsschwerpunkt auf Informationsvisualisierung liegt, folgendermaßen: „The overall theme of our work is to improve the experience of people using computers, making that experience more effective and enjoyable.“²¹⁴

²¹⁴ Bederson; Shneiderman (Verf. und Hrsg.): The Craft of Information Visualization, 2003, S. IX

7 LITERATURVERZEICHNIS

7.1 Literatur

- Assoziativ-OPAC Spidersearch von BOND. In: B.I.T.online 6 (2003) 3, S. 281
- Bederson, Benjamin; Shneiderman, Ben (Verf. und Hrsg.): The Craft of Information Visualization. Readings and Reflections. San Francisco: Morgan Kaufmann, 2003
- Beier, Heiko: Intelligente Informationsstrukturierung und TextMining mit semantischen Netzen. Beitrag zur 25. DGI-Online-Tagung Competence in Content. Online: http://www.agi-imc.de/isearch/is_dgi_articles.nsf/fsDocumentDispWeb?OpenFrameSet&Frame=Document&Src=%2Fisearch%2Fis_dgi_articles.nsf%2F0%2F4aabc7d7ae9b99a3c1256eae003b3072%3FOpenDocument%26Query%3Dsemantische%2520Netze%26AutoFramed%26Typ%3DA [Stand: 2003, Abrufdatum: 13.09.2004, Datei: DGI_2003_OT_Beier%202003.pdf]
- Beier, Heiko: Menschengerechte Aufbereitung von Information. Vom assoziativen Gedächtnis zum Semantischen Netz. Online: http://leuchttum.fh-koeln.de/uploads/media/Semantische_Netze.pdf [Stand: 2003, Abrufdatum: 22.11.2004, Datei: Semantische_Netze.pdf]
- BiblioMondo: ZONE-1 als neues Produkt. In: ABI-Technik 23 (2003) 1, S. 65
- BOND: Assoziativ-OPAC SpiderSearch. In: Bibliotheksdienst 37 (2003) 8/9, S. 1145
- BOND: Faszinierende Suchergebnisse durch visuell animierte Medienrecherche. In: ABI-Technik 23 (2003) 3, S. 247
- BOND: SpiderSearch – der neue grafische Assoziativ-OPAC. In: Bibliotheksdienst 36 (2002) 10, S. 1389f.
- „BOND-Web-OPAC der Stadtbibliothek Herten“. Online: <http://www.herten.de/kultur/bibliothek/index.htm> und Aufrufen des Links „Hier geht es zur Recherche im Online-Katalog der Stadtbibliothek Herten“
- Brockhaus multimedial 2005 premium. Mannheim: Bibliographisches Institut & F. A. Brockhaus, 2004
- „Broschüre zu SpiderSearch der Firma BOND“. Online: <http://www.bond-online.de/pdf/spidersearch.pdf> [Abrufdatum: 08.03.2005, Datei: spidersearch.pdf]
- „Broschüre zum System ZONES der Firma BiblioMondo“. Online: http://www.bibliomondo.com/data/ZONE-1_Brochure_EN.pdf [Abrufdatum: 05.10.2004, Datei: ZONE-1_Brochure_EN.pdf]
- Burkart, Margarete: Thesaurus. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, Handbuch zur Einführung in die Informationswissenschaft und -praxis. Hrsg. von Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar. 5., völlig neu gefasste Ausgabe. München: Saur, 2004, S. 141-154
- Card, Stuart; Mackinlay, Jock; Shneiderman, Ben (Verf. und Hrsg.): Readings in Information Visualization. Using Vision to Think. San Francisco: Morgan Kaufmann, 1999

-
- Däßler, Rolf: Informationsvisualisierung. Stand, Kritik und Perspektiven. Potsdam: Fachhochschule Potsdam, Fachbereich Archiv/Bibliothek/Dokumentation, Projektgruppe InfoViz. Online: <http://fabdp.fh-potsdam.de/daessler/paper/InfoVis99.pdf> [Stand: 1999, Abrufdatum: 30.09.2004, Datei: InfoVis99.pdf]
- Däßler, Rolf; Palm, Hartmut: Virtuelle Informationsräume mit VRML. Informationen recherchieren und präsentieren in 3D. 1. Aufl. Heidelberg: dpunkt, 1998
- Döring, Nicola: Computervermittelte Kommunikation, Mensch-Computer-Interaktion. In: Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis. Hrsg. von Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar. 5., völlig neu gefasste Ausgabe. München: Saur, 2004, S. 351-362
- Eibl, Maximilian: Visualisierung im Document Retrieval. Theoretische und praktische Zusammenführung von Softwareergonomie und Graphik Design. 2. überarbeitete und erweiterte Auflage. Bonn: Informationszentrum Sozialwissenschaften Bonn, 2003. (Forschungsberichte ; 7)
- Englmeier, Kurt: Ein Computer mit einem offenen Interface für Informationsprobleme. In: Mensch und Maschine. Informationelle Schnittstellen der Kommunikation. Hrsg. von Zimmermann, Harald; Luckhardt, Heinz-Dirk; Schulz, Angelika. Konstanz, Universitäts-Verlag, 1992, S. 229-244. (Schriften zur Informationswissenschaft ; 7)
- Ferber, Reginald: Information Retrieval. Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. 1. Aufl. Heidelberg: dpunkt, 2003
- Fuhr, Norbert: Information Retrieval. Skriptum zur Vorlesung im SS 04. Duisburg: Universität Duisburg-Essen, Campus Duisburg, Fachbereich Ingenieurwissenschaften, Institut für Informatik und Interaktive Systeme, Fachgebiet Informationssysteme. Online: http://www.is.informatik.uni-duisburg.de/courses/ir_ss04/folien/irskall.pdf [Stand: SS 2004, Abrufdatum: 19.01.2005, Datei: irskall.pdf]
- Fuhr, Norbert: Konzepte zur Gestaltung zukünftiger Information-Retrieval-Systeme. In: Experimentelles und praktisches Information Retrieval. Festschrift für Gerhard Lustig. Hrsg. von Rainer Kuhlen. Konstanz: Universitäts-Verlag Konstanz, 1992. (Schriften zu Informationswissenschaft ; 3), S. 59-75
- Gödert, Winfried: Navigation und Konzepte für ein interaktives Retrieval im OPAC. Oder: Von der Informationserschließung zur Wissenserkundung. In: Mitteilungen der VÖB 57 (2004) 1, S. 70-80. Online: http://voeb.uibk.ac.at/vm/vm57_1_goedert.pdf [Stand: 2004, Abrufdatum: 08.04.2005, Datei: vm57_1_goedert.pdf]
- Gödert, Winfried: Sacherschließung. Zweck, Aufgaben, Methoden: eine Vorstellung. In: Materialien zur Inhaltserschließung. Köln: Fachhochschule Köln, Institut für Informationswissenschaft, WS 2003/04. CD-ROM. [Datei: Sacherschliessung_Einführung.pdf]
- Gödert, Winfried: Sprache und Bedeutung. In: Materialien zur Inhaltserschließung. Köln: Fachhochschule Köln, Institut für Informationswissenschaft, WS 2003/04. CD-ROM. [Datei: Sprache_Bedeutung.pdf]
- Gödert, Winfried: Aufbereitung und Rezeption von Information. In: Medien- Informationsmanagement. Archivarische, dokumentarische, betriebswirtschaftliche, rechtliche und Berufsbild-Aspekte. Hrsg. von Marianne Englert; Lange, Eckhard; Schmitt, Heiner. Münster: LIT, 2003. (Beiträge zur Mediendokumentation ; 6), S. 139-152

-
- Gödert, Winfried: Der konstruktivistische Ansatz für Kommunikation und Informationsverarbeitung. In: Wolfenbütteler Notizen zur Buchgeschichte 27 (2002) 2, S. 199-218
- Gödert, Winfried; Lepsky, Klaus: Semantische Umfeldsuche im Information Retrieval in Online-Katalogen. Köln: Fachhochschule Köln, Fachbereich Bibliotheks- und Informationswesen, 1997. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft ; 7)
- Graph (Artikel). In: Brockhaus multimedial 2005 premium. Mannheim: Bibliographisches Institut & F. A. Brockhaus, 2004. [Datei (Screenshot): Graph_Brockhaus.jpg]
- Halbach, Wulf: Interfaces. Medien- und kommunikationstheoretische Elemente einer Interface-Theorie. München: Fink, 1994
- Hammwöhner, Rainer: Hypertext. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, Handbuch zur Einführung in die Informationswissenschaft und -praxis. Hrsg. von Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar. 5., völlig neu gefasste Ausgabe. München: Saur, 2004, S. 419-428
- Hauer, Manfred: Durch Content-Ergänzung, maschinelle Indexierung und modernes Information Retrieval können Recherchen in Bibliothekskatalogen deutlich verbessert werden. In: ABI-Technik 24 (2005) 4, S. 262-268
- Hauer, Manfred: Silicon Valley Vorarlberg: Maschinelle Indexierung und semantisches Retrieval verbessert den Katalog der Vorarlberger Landesbibliothek. In: Mitteilungen der VÖB 57 (2004) 3/4, S. 33-38
- Hauer, Manfred: Digitalisierung von Aufsätzen und anderen Texten mit maschineller Inhaltserschließung erschließt Wissensressourcen in Bibliotheken, Museen und Dokumentationszentren. Online: [http://www.agi-imc.de/internet.nsf/0/0fe6d2dd3fd0e16cc1256d4b0030b38f/\\$FILE/OWFTM_hauer2003capt.pdf](http://www.agi-imc.de/internet.nsf/0/0fe6d2dd3fd0e16cc1256d4b0030b38f/$FILE/OWFTM_hauer2003capt.pdf) [Stand: 2003, Abrufdatum: 31.08.2004, Datei: OWFTM_hauer2003capt.pdf]
- Hauer, Manfred: Neue Mathematik in Bibliotheken? 2 = 100. AGI öffnet semantische Search-Engine für wissenschaftliche Literatur: www.dandelon.com. Online: [http://www.agi-imc.de/iframe/FAQ_in_iS.nsf/0/0c8bfb37b1fc9c25c1256e5a005d19ff/\\$FILE/dandeloncom.pdf](http://www.agi-imc.de/iframe/FAQ_in_iS.nsf/0/0c8bfb37b1fc9c25c1256e5a005d19ff/$FILE/dandeloncom.pdf) [Abrufdatum: 31.08.2004, Datei: dandeloncom.pdf]
- Heinecke, Andreas: Mensch-Computer-Interaktion. München: Fachbuchverlag Leipzig im Carl Hanser Verlag, 2004
- „Hilfetext im Brockhaus multimedial 2005 premium – Wissensnetz“. [Datei (Screenshot): Hilfetext Brockhaus multimedial 2005 - Wissensnetz.jpg]
- „Homepage des Brockhaus multimedial – Produktdaten“. Online: <http://www.brockhaus-multimedial.de/daten/> [Abrufdatum: 10.03.2005, Datei: Brockhaus multimedial 2005.html]
- „Homepage der Firma AGI - Information Management Consultants“. Online: <http://www.agi-imc.de/> [Abrufdatum: 07.03.2005, Datei: www.agi-imc.de.html]
- „Homepage der Firma BiblioMondo“. Online: <http://www.bibliomondo.com/?language=de> [Abrufdatum: 07.03.2005, Datei: BiblioMondo1.html]
- „Homepage der Firma BiblioMondo – Kunden in Deutschland“. Online: <http://www.bibliomondo.com/site/clients/clients.php?subSection=0506> [Abrufdatum: 16.03.2005, Datei: clients.php.html]

-
- „Homepage der Firma BOND – zu SpiderSearch“. Online: <http://www.bond-online.de>
[Abrufdatum: 08.03.2005, Datei: Bond Bibliothekssysteme.html]
- „Homepage OLIVE“. Online: <http://www.otal.umd.edu/Olive/> [Stand: 1997, Abrufdatum:
27.09.2004, Datei: Olive.htm]
- „Homepage des Portals dandelon.com“. Online:
<http://www.dandelon.com/intelligentsearch.nsf/fmQSF?OpenForm> [Abrufdatum:
19.03.2005, Datei: fmQSF.html]
- „Homepage des Portals dandelon.com – Frequently Asked Questions“. Online:
http://www.agi-imc.de/icapture/FAQ_in_iS.nsf [Abrufdatum: 07.03.2005, Datei:
FAQ_in_iS.nsf.html]
- „Homepage Portal Informationswissenschaft“. Online: http://www.agi-imc.de/isearch/is_dgi.nsf
- „Homepage des Verbandes Bildung und Erziehung Landesverband NRW – Neuheiten des Brockhaus multimedial 2001“. Online: <http://www.vbe-nrw.de/vbe.html?ClusterRef=640>
[Abrufdatum: 23.10.2004, Datei: VBE - Verband Bildung und Erziehung, Landesverband NRW.htm]
- Interaktion (Artikel). In: Brockhaus multimedial 2005 premium. Mannheim: Bibliographisches Institut & F. A. Brockhaus, 2004. [Datei (Screenshot): Interaktion_Brockhaus.jpg]
- Kowalski, Gerald; Maybury, Mark: Information Storage and Retrieval Systems. Theory and Implementation. Second Edition. Boston: Kluwer, 2000. (The Kluwer International Series on Information Retrieval ; 8)
- Kuhlen, Rainer: Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank. Berlin: Springer, 1991
- Lepsky, Klaus: Im Heuhaufen suchen – und finden. Automatische Erschließung von Internetquellen: Möglichkeiten und Grenzen. In: BuB 50 (1998) 5, S.336-340
- Lepsky, Klaus: Vom OPAC zum Hyperkatalog. Daten und Indexierung. In: Erschließen, Suchen, Finden. Hrsg. von Hermes, Hans-Joachim; Wätjen, Hans-Joachim. Oldenburg: Bibliothek- und Informationssystem der Universität Oldenburg, 1996, S. 65-74
- Lepsky, Klaus: Inhaltserschließung II. In: Lehrmaterialien zur Veranstaltung. Vorlesungsskript. Köln: Fachhochschule Köln, Institut für Informationswissenschaft. Online: <http://www.iws.fh-koeln.de/institut/personen/lepsyk/material/lehre/Skript-1c-Grundlagen-der-Inhaltserschliessung-III.pdf> [Abrufdatum: 09.02.2005, Datei: Skript-1c-Grundlagen-der-Inhaltserschliessung-III.pdf]
- Lepsky, Klaus: Methoden und Verfahren des Information Retrieval. In: Lehrmaterialien zur Veranstaltung. Vorlesungsskript. Köln: Fachhochschule Köln, Institut für Informationswissenschaft. Online: <http://www.iws.fh-koeln.de/institut/personen/lepsyk/material/lehre/Skript-2-Methoden-und-Verfahren-des-Information-Retrieval.pdf> [Abrufdatum: 09.02.2005, Datei: Skript-2-Methoden-und-Verfahren-des-Information-Retrieval.pdf]
- Nohr, Holger: Theorie des Information Retrieval II. Automatische Indexierung. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, Handbuch zur Einführung in die Informationswissenschaft und -praxis. Hrsg. von Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar. 5., völlig neu gefasste Ausgabe. München: Saur, 2004, S. 215-225

Nohr, Holger: Maschinelle Indexierung. Skript zur Vorlesung im Fach Inhaltliche Erschließung an der Hochschule für Bibliotheks- und Informationswesen (HBI) Stuttgart. Online: <http://www.iuk.hdm-stuttgart.de/nohr/IWS/Mi/MI.html> [Stand: 1999, Abrufdatum: 30.03.2005, Datei: MI.html]

N.N. 2005a: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005b: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005c: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005d: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005e: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005f: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005g: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005h: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1 [Abruf: 11.11.2005]

N.N. 2005i: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1
[Abruf: 11.11.2005]

N.N. 2005j: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1
[Abruf: 11.11.2005]

N.N. 2005k: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1
[Abruf: 11.11.2005]

N.N. 2005l: URL: http://katalog.stbib-koeln.de/alswww4.dll/APS_AQUA?Style=Portal2&SubStyle=&Theme=&Lang=GER&ResponseEncoding=utf-8&SID=S11691133279761&Style=Portal2&Method=AquaSearch&BrowseAsHloc=-1
[Abruf: 11.11.2005]

N.N. 2005m: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005n: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005o: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005p: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005q: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005r: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005s: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005t: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005u: URL: <http://194.174.233.134/spider/index.asp?nextpage=&time=13:35:37>
[Abruf: 11.11.2005]

N.N. 2005v: URL: <http://194.174.233.134/spider/index.asp?nextpage=&time=13:35:37>
[Abruf: 11.11.2005]

N.N. 2005w: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

N.N. 2005x: URL: <http://194.174.233.134/spider/index.asp>
[Abruf: 11.11.2005]

-
- Preim, Bernhard: Entwicklung interaktiver Systeme. Grundlagen, Fallbeispiele und innovative Anwendungsfelder. Berlin: Springer, 1999
- Rädler, Karl: In Bibliothekskatalogen „googlen“. Integration von Inhaltsverzeichnissen, Volltexten und WEB-Ressourcen in Bibliothekskataloge. In: Bibliotheksdienst 38 (2004) 7/8, S. 927-939
- Regeln für den Schlagwortkatalog. RSWK. Hrsg. von der Konferenz für Regelwerksfragen beim Deutschen Bibliotheksinstitut. 3., überarbeitete und erweiterte Aufl. Berlin: Dbi, 1998
- Regeln für den Schlagwortkatalog. RSWK. Bearbeitet von der Kommission des Deutschen Bibliotheksinstituts für Sacherschließung. Berlin: Dbi, 1986
- Reimer, Ulrich: Wissensbasierte Verfahren der Organisation und Vermittlung von Information. In: Grundlagen der praktischen Information und Dokumentation, Bd. 1, Handbuch zur Einführung in die Informationswissenschaft und -praxis. Hrsg. von Kühlen, Rainer; Seeger, Thomas; Strauch, Dietmar. 5., völlig neu gefasste Ausgabe. München: Saur, 2004, S. 155-166
- Revolutionäre neue Suchvariante mit neuem grafischen Assoziativ-OPAC Spidersearch von BOND. In: B.I.T.online 5 (2002) 3, S. 264
- Sachse, Elisabeth; Liebig, Martina; Gödert, Winfried: Automatische Indexierung unter Einbeziehung semantischer Relationen. Ergebnisse des Retrievaltests zum MILOS II-Projekt. Köln: Fachhochschule Köln, Fachbereich Bibliotheks- und Informationswesen, 1998. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft ; 14)
- Salton, Gerard; McGill, Michael: Information Retrieval. Grundlegendes für Informationswissenschaftler. Hamburg: McGraw-Hill, 1987
- Schwartz, Dieter: Mehr Information durch Visualisierung von Daten? Konventionelle und innovative Visualisierungstechniken. In: B.I.T.online 2 (1999) 4, S. 455-462
- Shneiderman, Ben: User Interface Design. 3. Aufl. Bonn: mitp, 2002
- Sieber, Wolfram: Visualisierung von Thesaurusstrukturen unter besonderer Berücksichtigung eines Hyperbolic Tree Views. Diplomarbeit im Studiengang Informationswirtschaft, Fakultät für Informations- und Kommunikationswissenschaft, Fachhochschule Köln, 2004
- Spidersearch von BOND – Neue Suchvariante mit grafischem Assoziativ-OPAC. In: ABI-Technik 22 (2002) 3, S. 278
- Stadtbibliothek Reutlingen: Web-Auftritt komplett mit Web-OPAC Premium von BOND realisiert. In: BuB 57 (2005) 4, S. 271
- Stock, Wolfgang: Wissensrepräsentation. In: Lehre. Theoretische Informationswissenschaft. Düsseldorf: Heinrich-Heine-Universität Düsseldorf, Philosophische Fakultät, Institut für Sprache und Information, Abteilung Informationswissenschaft. Online: http://www.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/1/1111487239wissensrep.pdf [Stand: SS 2005, Abrufdatum: 08.04.2005, Datei: stock_wissensrep.pdf]
- Stock, Wolfgang: Einführung in die Informationswissenschaft. In: Lehre. Theoretische Informationswissenschaft. Düsseldorf: Heinrich-Heine-Universität Düsseldorf, Philosophische Fakultät, Institut für Sprache und Information, Abteilung Informationswissenschaft. Online: <http://www.phil-fak.uni->

-
- duesseldorf.de/infowiss/admin/public_dateien/files/1/1111487288einfuehrun.pdf [Stand: WS 2004/05, Abrufdatum: 08.04.2005, Datei: stock_einfuehrun.pdf]
- Stys, Thomas; Klos, Johanna; Kunkel, Paul: Thesauri entwickeln und anwenden mit MIDOSThesaurus. In: Information - Wissenschaft und Praxis 54 (2003) 5, S. 273-280
- „Suchhilfe zum Portal Informationswissenschaft und dandelon.com“. Online: http://www.agi-imc.de/isearch/is_dgi.nsf/helppage?OpenPage [Abrufdatum: 07.03.2005, Datei: helppage.html]
- Thissen, Frank: Kompendium Screen-Design. Effektiv informieren und kommunizieren mit Multimedia. 3. überarbeitete und erweiterte Auflage. Berlin: Springer, 2003
- Wersig, Gernot: Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. 2. ergänzte Auflage. München: Saur, 1985. (DGD-Schriftenreihe ; 8)
- Zeidler, Alfred; Zellner, Rudolf: Software-Ergonomie. Techniken der Dialoggestaltung. 2. verbesserte Auflage. München: Oldenbourg, 1994
- „ZONES-Web-OPAC der Stadtbibliothek Köln“. Online: https://katalog.stbib-koeln.de/alswww2.dll/APS_OPAC&Style=Portal2

7.2 Weitere Quellen

Unveröffentlicht

- „Anwenderbibliotheken von SpiderSearch nach Angaben der Firma BOND“. [Stand: 28.10.2004, Datei: Anwenderliste Spider Search.xls]
- „Vertrauliche Angaben zum Produkt WordSurfer der Firma BiblioMondo“ [Stand: 18.10.2004, Datei: WordSurfer.doc]
- Gödert, Winfried; Hartmann, Sarah; Hubrich, Jessica; Lepsky, Klaus; Schulenburg, Karin; Trunk, Daniela: Semantische Anreicherung der Schlagwortnormdatei. Ergebnisbericht. Köln: Fachhochschule Köln, Institut für Informationswissenschaft, 2004. [Datei: Ergebnisbericht.pdf]
- Sieber, Wolfram: Visualisierung von Thesaurusstrukturen unter besonderer Berücksichtigung eines Hyperbolic Tree Views. Diplomarbeit im Studiengang Informationswirtschaft. Köln: Fachhochschule Köln, Fakultät für Informations- und Kommunikationswissenschaft, 2004

E-Mail

- „E-Mail Brockhaus“ [Datum: 26.11.2004, Verfasser: Arnd Frederichs, technischer Projektleiter des Brockhaus multimedial, Datei: Antwort_Diplomarbeit.eml]

Daten

- „Die Schlagwortnormdatei aus dem Bereich Bibliotheks- und Archivwesen im BasisPlus-Format“. [Verzeichnis: SWD_BasisPlus, Datei: SWD_Bib.txt und SWD_Nachrelationierung_Bib.txt]